THINKING, FAST AND SLOW

# DANIEL KAHNEMAN

# SCHNELLES DENKEN LANGSAMES DENKEN

DER SPIEGEL BESTSELLER

NOBELPREIS FÜR WIRTSCHAFT

»Daniel Kahneman ist ganz sicher der wichtigste Psychologe unserer Zeit. Das Erscheinen dieses Buchs ist ein Großereignis.«
STEVEN PINKER

Siedler

*Zum Andenken an Amos Tversky*

# Inhaltsverzeichnis

*Widmung*
*Einleitung*
*Wo wir heute stehen*
*Was als Nächstes kommt*
**TEIL I - Zwei Systeme**
1. Die Figuren der Geschichte
*Zwei Systeme*
*Der Gang der Handlung – ein kurzer Überblick*
*Konflikt*
*Illusionen*
*Nützliche Fiktionen*
2. Aufmerksamkeit und Anstrengung
*Mentale Anstrengung*
3. Der faule Kontrolleur
*Das ausgelastete und erschöpfte System 2*
*Das faule System 2*
*Intelligenz, Kontrolle und Rationalität*
4. Die Assoziationsmaschine
*Die Wunder des Priming*
*Primes, die uns anleiten*
5. Kognitive Leichtigkeit
*Illusionen des Gedächtnisses*
*Illusionen der Wahrheit*
*Wie man eine überzeugende Mitteilung schreibt*
*Beanspruchung und Anstrengung*
*Die Freuden mühelosen Denkens*
*Leichtigkeit, Stimmung und Intuition*
6. Normen, Überraschungen und Ursachen
*Normalität beurteilen*
*Ursachen und Intentionen*
7. Eine Maschine für voreilige Schlussfolgerungen

*Vernachlässigung von Ambiguität und Unterdrückung von Zweifeln*
*Die Vorliebe, Aussagen zu glauben und eigene Erwartungen zu bestätigen*
*Überzogene emotionale Kohärenz – der Halo-Effekt*
*What you see is all there is*
8. Wie wir Urteile bilden
*Elementare Bewertungen*
*Mengen und Prototypen*
*Intensitäten und wie man sie vergleichen kann*
*Die mentale Schrotflinte*
9. Eine leichtere Frage beantworten
*Fragen ersetzen*
*Die 3-D-Heuristik*
*Die Stimmungsheuristik für Glück*
*Die Affektheuristik*
**TEIL II - Heuristiken und kognitive Verzerrungen**
10. Das Gesetz der kleinen Zahlen
*Das Gesetz der kleinen Zahlen*
*Die Tendenz, eher zu glauben als zu zweifeln*
*Ursache und Zufall*
11. Anker
*Ankerung als Anpassung*
*Ankerung als ein Priming-Effekt*
*Der Ankerungsindex*
*Gebrauch und Missbrauch von Ankern*
*Ankerung und die beiden Systeme*
12. Die Wissenschaft der Verfügbarkeit
*Die Psychologie der Verfügbarkeit*
13. Verfügbarkeit, Emotion und Risiko
*Verfügbarkeit und Affekt*
*Die Öffentlichkeit und die Experten*
14. Was studiert Tom W.?
*Vorhersage durch Repräsentativität*
*Die Sünden der Repräsentativität*
*Wie man die Intuition diszipliniert*
15. Linda: Weniger ist mehr
*Weniger ist mehr, manchmal sogar bei gemeinsamer Bewertung*
16. Ursachen vs. Statistik

*Kausale Stereotype*
*Kausale Situationen*
*Kann man Psychologie unterrichten?*
17. Regression zum Mittelwert
*Talent und Glück*
*Regression verstehen*
18. Intuitive Vorhersagen bändigen
*Nicht regressive Intuitionen*
*Eine Korrektur für intuitive Vorhersagen*
*Eine Verteidigung extremer Vorhersagen?*
*Die Regression im Zwei-Systeme-Modell*
**TEIL III - Selbstüberschätzung**
19. Die Illusion des Verstehens
*Die sozialen Kosten der Rückschau*
*Erfolgsrezepte*
20. Die Illusion der Gültigkeit
*Die Illusion der Gültigkeit*
*Der Irrglaube, einen guten Riecher für Aktien zu haben*
*Kompetenz und Gültigkeit – warum wir diesen Illusionen unterliegen*
*Die Illusionen von Experten*
*Die Experten können nichts dafür – die Welt ist eben kompliziert*
21. Intuitionen und Formeln
*Häufigkeit des Geschlechtsverkehrs minus Streithäufigkeit*
*Die Feindseligkeit gegen Algorithmen*
*Von Paul Meehl lernen*
*Probieren Sie es selbst aus*
22. Die Intuition von Experten: Wann können wir ihr vertrauen?
*Glanzleistungen und Mängel*
*Intuition als Wiedererkennen*
*Erwerb von Fertigkeiten*
*Die geeignete Umgebung für Expertise*
*Feedback und Übung*
*Die Gültigkeit von Intuitionen beurteilen*
23. Die Außensicht
*Die Verlockung der Innensicht*
*Der Planungsfehlschluss*
*Wie man Planungsfehlschlüsse in den Griff bekommt*

*Entscheidungen und Irrtümer*
*Einen Test nicht bestehen*
24. Die Maschine des Kapitalismus
*Optimisten*
*Illusionen von Unternehmern*
*Vernachlässigung der Konkurrenz*
*Selbstüberschätzung*
*Wie die Prä-mortem-Methode helfen kann*
**TEIL IV - Entscheidungen**
25. Irrtümer
*Bernoullis Irrtum*
26. Die Neue Erwartungstheorie
*Verlustaversion*
*Blinde Flecken der Neuen Erwartungstheorie*
27. Der Endowment-Effekt
*Der Endowment-Effekt*
*Wie ein Wertpapierhändler denken*
28. Negative Ereignisse
*Ziele sind Referenzpunkte*
*Den Status quo verteidigen*
*Das Rechtswesen und die Verlustaversion*
29. Das viergeteilte Muster
*Wahrscheinlichkeiten verändern*
*Das Allais-Paradoxon*
*Entscheidungsgewichte*
*Das viergeteilte Muster*
*Glücksspiele im Schatten des Gesetzes*
30. Seltene Ereignisse
*Überschätzen und Übergewichten*
*Anschauliche Ergebnisse*
*Anschauliche Wahrscheinlichkeiten*
*Entscheidungen auf der Basis globaler Eindrücke*
31. Risikostrategien
*Weit oder eng?*
*Samuelsons Problem*
*Risikostrategien*
32. Buch führen

*Mentale Buchführung*
*Reue*
*Verantwortung*
33. Umkehrungen
*Eine Herausforderung für die Ökonomik*
*Kategorien*
*Ungerechte Umkehrungen*
34. Frames und Wirklichkeit
*Emotionales Framing*
*Leere Intuitionen*
*Gute Frames*
**TEIL V - Zwei Selbste**
35. Zwei Selbste
*Erfahrungsnutzen*
*Erfahrung und Gedächtnis*
*Welches Selbst sollte zählen?*
*Biologie kontra Rationalität*
36. Das Leben als eine Geschichte
*Amnestischer Urlaub*
37. Erlebtes Wohlbefinden
*Erlebtes Wohlbefinden*
38. Lebenszufriedenheit
*Die Fokussierungs-Illusion*
*Immer wieder der Faktor Zeit*
*Schlusswort*
*Zwei Selbste*
*Econs und Humans*
*Zwei Systeme*
*ANHANG*
*Urteile unter Unsicherheit: Heuristiken und kognitive Verzerrungen*
*Repräsentativität*
*Verfügbarkeit*
*Anpassung und Verankerung*
*Diskussion*
*Zusammenfassung*
*Entscheidungen, Werte und Frames*

*Risikobehaftete Entscheidungen*
*Framing der Ergebnisse*
*Die Psychophysik von Wahrscheinlichkeiten*
*Formulierungseffekte*
*Transaktionen und Tauschgeschäfte*
*Verluste und Kosten*
*Abschließende Bemerkungen*
*Dank*
*Anmerkungen*
*TEIL I*
*1. Die Figuren der Geschichte*
*2. Aufmerksamkeit und Anstrengung*
*3. Der faule Kontrolleur*
*4. Die Assoziationsmaschine*
*5. Kognitive Leichtigkeit*
*6. Normen, Überraschungen und Ursachen*
*7. Eine Maschine für voreilige Schlussfolgerungen*
*8. Wie wir Urteile bilden*
*9. Eine leichtere Frage beantworten*
*TEIL II*
*10. Das Gesetz der kleinen Zahlen*
*11. Anker*
*12. Die Wissenschaft der Verfügbarkeit*
*13. Verfügbarkeit, Emotion und Risiko*
*14. Was studiert Tom W.?*
*15. Linda: Weniger ist mehr*
*16. Ursachen vs. Statistik*
*17. Regression zum Mittelwert*
*18. Intuitive Vorhersagen bändigen*
*TEIL III*
*19. Die Illusion des Verstehens*
*20. Die Illusion der Gültigkeit*
*21. Intuitionen und Formeln*
*22. Die Intuition von Experten: Wann können wir ihr vertrauen?*
*23. Die Außensicht*
*24. Die Maschine des Kapitalismus*
*TEIL IV*

*25. Irrtümer*
*26. Die Neue Erwartungstheorie*
*27. Der Endowment-Effekt*
*28. Negative Ereignisse*
*29. Das viergeteilte Muster*
*30. Seltene Ereignisse*
*31. Risikostrategien*
*32. Buch führen*
*33. Umkehrungen*
*34. Frames und Wirklichkeit*
*TEILV*
*35. Zwei Selbste*
*36. Das Leben als eine Geschichte*
*37. Erlebtes Wohlbefinden*
*38. Lebenszufriedenheit*
*Schlusswort*
*ANHANG Urteile unter Unsicherheit*
*Entscheidungen, Werte und Frames*
*Sachregister*
*Personenregister*
*Copyright*

# *Einleitung*

Jedem Autor, vermute ich mal, schwebt eine Situation vor, in der Leser seines Werks von der Lektüre desselben profitieren könnten. Ich denke dabei an den Kaffeeautomaten im Büro, vor dem Mitarbeiter Ansichten und Tratsch miteinander austauschen. Meine Hoffnung ist, dass ich den Wortschatz bereichere, den Menschen benutzen, wenn sie sich über Urteile und Entscheidungen anderer, die neue Geschäftsstrategie ihres Unternehmens oder die Anlageentscheidungen eines Kollegen unterhalten. Weshalb sich mit Tratsch befassen? Weil es viel leichter und auch viel angenehmer ist, die Fehler anderer zu erkennen und zu benennen als seine eigenen. Selbst unter den günstigsten Umständen fällt es uns schwer, unsere Überzeugungen und Wünsche zu hinterfragen, und es fällt uns besonders schwer, wenn es am nötigsten wäre – aber wir können von den sachlich fundierten Meinungen anderer profitieren. Viele von uns nehmen in Gedanken von sich aus vorweg, wie Freunde und Kollegen unsere Entscheidungen beurteilen werden; deshalb kommt es maßgeblich auf Qualität und Inhalt dieser vorweggenommenen Urteile an. Die Erwartung intelligenten Geredes über uns ist ein starkes Motiv für ernsthafte Selbstkritik, stärker als alle an Silvester gefassten guten Vorsätze, die Entscheidungsfindung am Arbeitsplatz und zu Hause zu verbessern.

Um zuverlässige Diagnosen zu stellen, muss ein Arzt eine Vielzahl von Krankheitsbezeichnungen lernen, und jeder dieser Termini verknüpft ein Konzept der Erkrankung mit ihren Symptomen, möglichen Vorstufen und Ursachen, möglichen Verläufen und Konsequenzen sowie möglichen Eingriffen zur Heilung oder Linderung der Krankheit. Das Erlernen der ärztlichen Heilkunst besteht auch darin, die medizinische Fachsprache zu erlernen. Um Urteile und Entscheidungen besser verstehen zu können, bedarf es eines reichhaltigeren Wortschatzes, als ihn die Alltagssprache zur Verfügung stellt. Die Tatsache, dass unsere Fehler charakteristische Muster aufweisen, begründet die Hoffnung darauf, dass andere in sachlich fundierter Weise über uns reden mögen. Systematische Fehler – auch »Verzerrungen« *(biases)* genannt – treten in vorhersehbarer Weise unter bestimmten Umständen auf. Wenn ein attraktiver und selbstbewusster

Redner dynamisch aufs Podium springt, kann man davon ausgehen, dass das Publikum seine Äußerungen günstiger beurteilt, als er es eigentlich verdient. Die Verfügbarkeit eines diagnostischen Etiketts für diesen systematischen Fehler – der Halo-Effekt – erleichtert es, ihn vorwegzunehmen, zu erkennen und zu verstehen.

Wenn Sie gefragt werden, woran Sie gerade denken, können Sie diese Frage normalerweise beantworten. Sie glauben zu wissen, was in Ihrem Kopf vor sich geht – oftmals führt ein bewusster Gedanke in wohlgeordneter Weise zum nächsten. Aber das ist nicht die einzige Art und Weise, wie unser Denkvermögen *(mind)* funktioniert, es ist nicht einmal seine typische Funktionsweise. Die meisten Eindrücke und Gedanken tauchen in unserem Bewusstsein auf, ohne dass wir wüssten, wie sie dorthin gelangten. Sie können nicht rekonstruieren, wie Sie zu der Überzeugung gelangten, eine Lampe stehe auf dem Schreibtisch vor Ihnen, wie es kam, dass Sie eine Spur von Verärgerung aus der Stimme Ihres Gatten am Telefon heraushörten, oder wie es Ihnen gelang, einer Gefahr auf der Straße auszuweichen, ehe Sie sich ihrer bewusst wurden. Die mentale Arbeit, die Eindrücke, Intuitionen und viele Entscheidungen hervorbringt, vollzieht sich im Stillen in unserem Geist.

Ein Schwerpunkt dieses Buches sind Fehler in unserem intuitiven Denken. Doch die Konzentration auf diese Fehler bedeutet keine Herabsetzung der menschlichen Intelligenz, ebenso wenig, wie das Interesse an Krankheiten in medizinischen Texten Gesundheit verleugnet. Die meisten von uns sind die meiste Zeit ihres Lebens gesund, und die meisten unserer Urteile und Handlungen sind meistens angemessen. Auf unserem Weg durchs Leben lassen wir uns normalerweise von Eindrücken und Gefühlen leiten, und das Vertrauen, das wir in unsere intuitiven Überzeugungen und Präferenzen setzen, ist in der Regel gerechtfertigt. Aber nicht immer. Wir sind oft selbst dann von ihrer Richtigkeit überzeugt, wenn wir irren, und ein objektiver Beobachter erkennt unsere Fehler mit höherer Wahrscheinlichkeit als wir selbst.

Und so wünsche ich mir, dass dieses Buch die Gespräche am Kaffeeautomaten dadurch verändert, dass es unsere Fähigkeit verbessert, Urteils- und Entscheidungsfehler von anderen und schließlich auch von uns selbst zu erkennen und verstehen, indem es dem Leser eine differenzierte und exakte Sprache an die Hand gibt, in der sich diese Fehler diskutieren lassen. Eine zutreffende Diagnose mag wenigstens in einigen Fällen eine

Korrektur ermöglichen, um den Schaden, den Fehlurteile und -entscheidungen verursachen, zu begrenzen.

Dieses Buch stellt mein gegenwärtiges Verständnis von Urteils- und Entscheidungsprozessen dar, das maßgeblich von psychologischen Entdeckungen der letzten Jahrzehnte geprägt wurde. Die zentralen Ideen gehen allerdings auf jenen glücklichen Tag des Jahres 1969 zurück, an dem ich einen Kollegen bat, als Gastredner in einem Seminar zu sprechen, das ich am Fachbereich Psychologie der Hebräischen Universität von Jerusalem hielt. Amos Tversky galt als ein aufstrebender Star auf dem Gebiet der Entscheidungsforschung – ja, auf allen Forschungsfeldern, auf denen er sich tummelte –, sodass ich wusste, dass es eine interessante Veranstaltung werden würde. Viele Menschen, die Amos kannten, hielten ihn für die intelligenteste Person, der sie je begegnet waren. Er war brillant, redegewandt und charismatisch. Er war auch mit einem vollkommenen Gedächtnis für Witze gesegnet und mit einer außergewöhnlichen Fähigkeit, mit ihrer Hilfe ein Argument zu verdeutlichen. In Amos' Gegenwart war es nie langweilig. Er war damals 32, ich war 35. Amos berichtete den Seminarteilnehmern von einem aktuellen Forschungsprogramm an der Universität Michigan, bei dem es um die Beantwortung der folgenden Frage ging: Sind Menschen gute intuitive Statistiker? Wir wussten bereits, dass Menschen gute intuitive Grammatiker sind: Ein vierjähriges Kind befolgt, wenn es spricht, mühelos die Regeln der Grammatik, obwohl es die Regeln als solche nicht kennt. Haben Menschen ein ähnlich intuitives Gespür für die grundlegenden Prinzipien der Statistik? Amos berichtete, die Antwort darauf sei ein bedingtes Ja. Wir hatten im Seminar eine lebhafte Diskussion, und wir verständigten uns schließlich darauf, dass ein bedingtes Nein eine bessere Antwort sei. Amos und mir machte dieser Meinungsaustausch großen Spaß, und wir gelangten zu dem Schluss, dass intuitive Statistik ein interessantes Forschungsgebiet sei und dass es uns reizen würde, dieses Feld gemeinsam zu erforschen. An jenem Freitag trafen wir uns zum Mittagessen im *Café Rimon*, dem Stammlokal von Künstlern und Professoren in Jerusalem, und planten eine Studie über die statistischen Intuitionen von Wissenschaftlern. Wir waren in diesem Seminar zu dem Schluss gelangt, dass unsere eigene Intuition unzureichend war. Obwohl wir beide schon jahrelang Statistik lehrten und anwandten, hatten wir kein intuitives Gespür für die Zuverlässigkeit statistischer Ergebnisse bei kleinen Stichproben entwickelt. Unsere subjektiven Urteile waren verzerrt: Wir

schenkten allzu bereitwillig Forschungsergebnissen Glauben, die auf unzureichender Datengrundlage basierten, und neigten dazu, bei unseren eigenen Forschungsarbeiten zu wenig Beobachtungsdaten zu erheben. [1] Mit unserer Studie wollten wir herausfinden, ob andere Forscher an der gleichen Schwäche litten.

Wir bereiteten eine Umfrage vor, die realistische Szenarien statistischer Probleme beinhaltete, die in der Forschung auftreten. Amos trug die Antworten einer Gruppe von Experten zusammen, die an einer Tagung der Society of Mathematical Psychology teilnahmen, darunter waren auch die Verfasser zweier Statistik-Lehrbücher. Wie erwartet fanden wir heraus, dass unsere Fachkollegen, genauso wie wir, die Wahrscheinlichkeit, dass das ursprüngliche Ergebnis eines Experiments auch bei einer kleinen Stichprobe erfolgreich reproduziert werden würde, enorm überschätzten. Auch gaben sie einer fiktiven Studentin sehr ungenaue Auskünfte über die Anzahl der Beobachtungsdaten, die sie erheben müsse, um zu einer gültigen Schlussfolgerung zu gelangen. Selbst Statistiker waren also keine guten intuitiven Statistiker. Als wir den Artikel schrieben, in dem wir diese Ergebnisse darlegten, stellten Amos und ich fest, dass uns die Zusammenarbeit großen Spaß machte. Amos war immer sehr witzig, und in seiner Gegenwart wurde auch ich witzig, sodass wir Stunden gewissenhafter Arbeit in fortwährender Erheiterung verbrachten. Die Freude, die wir aus unserer Zusammenarbeit zogen, machte uns ungewöhnlich geduldig; man strebt viel eher nach Perfektion, wenn man sich nicht langweilt. Am wichtigsten war vielleicht, dass wir unsere kritischen Waffen an der Tür abgaben. Sowohl Amos als auch ich waren kritisch und streitlustig – er noch mehr als ich, aber in den Jahren unserer Zusammenarbeit hat keiner von uns beiden irgendetwas, was der andere sagte, rundweg abgelehnt. Eine der größten Freuden, die mir die Zusammenarbeit mit Amos schenkte, bestand gerade darin, dass er viel deutlicher als ich selbst sah, worauf ich mit meinen vagen Gedanken hinauswollte. Amos war der bessere Logiker von uns beiden, er war theoretisch versierter und hatte einen untrüglichen Orientierungssinn. Ich hatte einen intuitiveren Zugang und war stärker in der Wahrnehmungspsychologie verwurzelt, aus der wir viele Ideen übernahmen. Wir waren einander hinreichend ähnlich, um uns mühelos zu verständigen, und wir waren hinreichend unterschiedlich, um uns gegenseitig zu überraschen. Wir verbrachten routinemäßig einen Großteil

unserer Arbeitstage zusammen, oftmals auf langen Spaziergängen. Während der kommenden 14 Jahre bildete diese Zusammenarbeit den Mittelpunkt unserer Leben, und unsere gemeinsamen Arbeiten aus dieser Zeit waren das Beste, was jeder von uns überhaupt an wissenschaftlichen Beiträgen lieferte.

Wir entwickelten schon bald eine bestimmte Vorgehensweise, die wir viele Jahre lang beibehielten. Unsere Forschung bestand in einem Gespräch, in dem wir Fragen erfanden und gemeinsam unsere intuitiven Antworten überprüften. Jede Frage war ein kleines Experiment, und wir führten an jedem Tag viele Experimente durch. Wir suchten nicht ernsthaft nach der richtigen Antwort auf die statistischen Fragen, die wir stellten. Wir wollten die intuitive Antwort herausfinden und analysieren, die erste, die uns einfiel, diejenige, die wir spontan geben wollten, auch wenn wir wussten, dass sie falsch war. Wir glaubten – richtigerweise, wie sich zeigte –, dass jede Intuition, die wir beide teilten, auch von vielen anderen geteilt würde und dass es leicht wäre, ihre Auswirkungen auf Urteile nachzuweisen. Einmal entdeckten wir zu unserer großen Freude, dass wir die gleichen verrückten Ideen über die zukünftigen Berufe mehrerer Kleinkinder hatten, die wir beide kannten. Wir identifizierten den streitlustigen dreijährigen Anwalt, den schrulligen Professor, den empathischen und leicht zudringlichen Psychotherapeuten. Natürlich waren diese Vorhersagen absurd, aber wir fanden sie trotzdem reizvoll. Es war auch klar, dass unsere Intuitionen von der Ähnlichkeit beeinflusst wurden, die das jeweilige Kind mit dem kulturellen Stereotyp eines bestimmten Berufs aufwies. Diese lustige Übung half uns dabei, eine Theorie zu entwickeln, die damals in unseren Köpfen im Entstehen begriffen war, und zwar über die Bedeutung der Ähnlichkeit bei Vorhersagen. Wir überprüften und verfeinerten diese Theorie in Dutzenden von Experimenten, wie etwa dem folgenden.

Wenn Sie über die nächste Frage nachdenken, sollten Sie davon ausgehen, dass Steve zufällig aus einer repräsentativen Stichprobe ausgewählt wurde:

> Eine Person wurde von einem Nachbarn wie folgt beschrieben: »Steve ist sehr scheu und verschlossen, immer hilfsbereit, aber kaum an anderen oder an der Wirklichkeit interessiert. Als sanftmütiger und ordentlicher Mensch hat er ein Bedürfnis

nach Ordnung und Struktur und eine Passion für Details.« Ist Steve eher Bibliothekar oder eher Landwirt?

Die Ähnlichkeit von Steves Persönlichkeit mit dem Stereotyp eines Bibliothekars fällt jedem sofort ins Auge, aber dabei werden sachdienliche statistische Erwägungen fast immer außer Betracht gelassen. Wussten Sie, dass in den Vereinigten Staaten auf jeden männlichen Bibliothekar zwanzig Landwirte kommen? Weil es so viel mehr Landwirte gibt, wird man höchstwahrscheinlich auch mehr »sanftmütige und ordentliche« Menschen auf Traktoren als hinter den Informationsschaltern von Bibliotheken finden. Wir stellten jedoch fest, dass diejenigen, die an unseren Experimenten teilnahmen, die relevanten statistischen Fakten außer Acht ließen und sich ausschließlich auf die Ähnlichkeit stützten. Wir nahmen an, dass sie die Ähnlichkeit als eine vereinfachende Heuristik (grob gesagt: eine Faustregel) benutzten, um ein schwieriges Urteil zu fällen. Das Vertrauen auf die Heuristik verursachte vorhersehbare Verzerrungen in ihren Vorhersagen.

Ein anderes Mal fragten Amos und ich uns nach der Scheidungsrate bei Professoren an unserer Universität. Wir bemerkten, dass die Frage eine Gedächtnissuche nach geschiedenen Professoren, die wir kannten oder von denen wir gehört hatten, auslöste und dass wir die Größe der Kategorien nach der Leichtigkeit beurteilten, mit denen uns die Beispiele einfielen. Wir nannten diese Orientierung an der Leichtigkeit der Gedächtnissuche »Verfügbarkeitsheuristik« *(availability heuristic)*. In einer unserer Studien forderten wir die Teilnehmer auf, eine einfache Frage über Wörter in einem typischen englischen Text zu beantworten:[2]

Betrachten Sie den Buchstaben K.
Taucht K eher als erster oder als dritter Buchstabe in einem Wort auf?

Wie jeder Scrabble-Spieler weiß, ist es viel leichter, Wörter zu finden, die mit einem bestimmten Buchstaben anfangen, als Wörter, die denselben Buchstaben an der dritten Stelle aufweisen. Dies gilt für alle Buchstaben

des Alphabets. Aus diesem Grund erwarteten wir, dass die Versuchsteilnehmer die Häufigkeit von Buchstaben an der ersten Stelle überschätzen würden – selbst der Buchstaben (wie K, L, N, R, V), die im Englischen tatsächlich häufiger an der dritten Stelle vorkommen. Hier führt der Rückgriff auf eine Heuristik erneut zu einem vorhersagbaren Urteilsfehler. So bezweifle ich beispielsweise mittlerweile meinen lange gehegten Eindruck, Ehebruch komme bei Politikern häufiger vor als bei Ärzten oder Anwälten. Ich hatte mir sogar Erklärungen für die »Tatsache« einfallen lassen, wie etwa die aphrodisierende Wirkung von Macht und die Verlockungen des Lebens fern von zu Hause. Ich erkannte schließlich, dass über die Fehltritte von Politikern mit viel höherer Wahrscheinlichkeit berichtet wird als über die von Anwälten und Ärzten. Mein intuitiver Eindruck war womöglich ausschließlich auf die Themenwahl von Journalisten und meine Anwendung der Verfügbarkeitsheuristik zurückzuführen.

Amos und ich verbrachten mehrere Jahre damit, Verzerrungen des intuitiven Denkens bei verschiedenen Aufgaben zu erforschen und zu dokumentieren: bei der Zuschreibung von Eintrittswahrscheinlichkeiten zu Ereignissen, bei Vorhersagen über zukünftige Ereignisse, bei der Beurteilung von Hypothesen und der Abschätzung von Häufigkeiten. Im fünften Jahr unserer Zusammenarbeit stellten wir unsere wichtigsten Ergebnisse im Wissenschaftsmagazin *Science* vor, das von Wissenschaftlern vieler Disziplinen gelesen wird. Der Artikel (der in ganzer Länge am Ende dieses Buches abgedruckt ist) trug den Titel »Judgement under Uncertainty: Heuristics and Biases« (»Urteile unter Unsicherheit: Heuristiken und kognitive Verzerrungen«). Er beschrieb die vereinfachenden Abkürzungen des intuitiven Denkens und erläuterte etwa zwanzig kognitive Verzerrungen als Beispiele dieser Heuristiken – und auch, um die Rolle dieser Heuristiken bei der Urteilsbildung zu verdeutlichen.

Wissenschaftshistoriker haben oft darauf hingewiesen, dass bei den Wissenschaftlern eines Fachgebiets zu jedem beliebigen Zeitpunkt Einigkeit über die Grundannahmen ihrer Disziplin besteht. Sozialwissenschaftler bilden da keine Ausnahme; sie gehen von einem Bild der menschlichen Natur aus, das den Hintergrund der meisten Diskussionen über bestimmte Verhaltensweisen bildet und nur selten infrage gestellt wird. In den 1970er-Jahren erachteten die meisten Sozialwissenschaftler zwei Annahmen über die menschliche Natur als erwiesen. Zum einen, dass Menschen sich im

Allgemeinen rational verhalten und normalerweise klar denken. Zum anderen, dass Emotionen wie Furcht, Zuneigung und Hass die meisten Fälle erklären, in denen Menschen von der Rationalität abweichen.

Unser Artikel stellte beide Annahmen infrage, ohne sie direkt zu diskutieren. Wir dokumentierten systematische Fehler im Denken gewöhnlicher Menschen und führten diese Fehler auf die Konstruktion des Kognitionsmechanismus zurück und nicht auf die Verfälschung des Denkens durch Emotionen.

Unser Artikel stieß auf viel mehr Beachtung, als wir erwartet hatten, und er ist noch immer einer der meistzitierten Beiträge in den Sozialwissenschaften (im Jahr 2010 bezogen sich über 300 wissenschaftliche Aufsätze darauf). Wissenschaftler in anderen Disziplinen fanden ihn nützlich, und das Konzept der Heuristiken und kognitiven Verzerrungen wurde in vielen Bereichen, wie etwa bei der medizinischen Diagnose, der Rechtsprechung, der Auswertung nachrichtendienstlicher Erkenntnisse, der Philosophie, der Finanzwissenschaft, der Statistik und der Militärstrategie, gewinnbringend angewendet.

So haben beispielsweise Politikwissenschaftler herausgefunden, dass die Verfügbarkeitsheuristik zu erklären hilft, weshalb einige Probleme in der Öffentlichkeit große Aufmerksamkeit finden, während andere vernachlässigt werden. Menschen neigen dazu, die relative Bedeutung von Problemen danach zu beurteilen, wie leicht sie sich aus dem Gedächtnis abrufen lassen – und diese Abrufleichtigkeit wird weitgehend von dem Ausmaß der Medienberichterstattung bestimmt. Häufig erwähnte Themen ziehen unsere Aufmerksamkeit auf sich, während andere aus dem Bewusstsein verschwinden. Andererseits entspricht das, worüber die Medien berichten, ihrer Einschätzung dessen, was die Öffentlichkeit gegenwärtig bewegt. Es ist kein Zufall, dass autoritäre Regime unabhängige Medien unter erheblichen Druck setzen. Da das Interesse der Öffentlichkeit am leichtesten durch dramatische Ereignisse und Stars geweckt wird, sind mediale »Fressorgien« weitverbreitet. So war es beispielsweise nach Michael Jacksons Tod mehrere Wochen lang praktisch unmöglich, einen Fernsehsender zu finden, der über ein anderes Thema berichtete. Über höchst wichtige, aber langweilige Themen, die weniger herzergreifend sind, wie etwa sinkende Bildungsstandards oder die Überinvestition medizinischer Ressourcen im letzten Lebensjahr, wird dagegen kaum berichtet. (Während ich dies schreibe, fällt mir auf, dass meine Auswahl

von Beispielen mit »geringer Medienpräsenz« von der Verfügbarkeit bestimmt wurde. Die Themen, die ich als Beispiele auswähle, werden häufig erwähnt; genauso wichtige Anliegen, die weniger verfügbar sind, fielen mir nicht ein.)

Damals war es uns nicht völlig klar, aber ein wesentlicher Grund für den breiten Anklang, den Heuristiken und kognitive Verzerrungen außerhalb der Psychologie finden, war ein zufälliges Merkmal unserer Arbeit: Wir nahmen fast immer den vollständigen Wortlaut der Fragen, die wir uns selbst und unseren Versuchspersonen stellten, in unsere Artikel auf. Diese Fragen dienten als Demonstrationen für den Leser; sie sollten ihm verdeutlichen, wie er selbst durch kognitive Verzerrungen zu Fehlurteilen gelangt. Ich hoffe, Sie hatten eine solche Erfahrung, als Sie die Frage über Steve, den Bibliothekar, lasen, die Ihnen helfen sollte, zu erkennen, wie stark Ähnlichkeit (mit Stereotypen) Wahrscheinlichkeitsaussagen beeinflusst und wie leicht relevante statistische Tatsachen ignoriert werden.

Die Fallbeispiele boten Wissenschaftlern unterschiedlichster Disziplinen – insbesondere aber Philosophen und Ökonomen – eine ungewöhnliche Gelegenheit, mögliche Fehler in ihrem eigenen Denken zu beobachten. Nachdem sie sich bei ihren eigenen Fehlern ertappt hatten, waren sie eher bereit, die damals vorherrschende dogmatische Annahme, jeder Mensch denke rational und logisch, infrage zu stellen. Die Wahl der Methode war entscheidend: Wenn wir nur über Ergebnisse konventioneller Experimente berichtet hätten, hätte der Artikel viel weniger Beachtung gefunden. Außerdem hätten sich skeptische Leser von den Ergebnissen distanziert, indem sie Urteilsfehler auf die bekannte Schludrigkeit von Studenten, die typischen Probanden psychologischer Studien, zurückgeführt hätten. Selbstverständlich haben wir Demonstrationsbeispiele nicht deshalb Standardexperimenten vorgezogen, weil wir Philosophen und Ökonomen beeinflussen wollten. Wir zogen Demonstrationen vor, weil sie unterhaltsamer waren, und wir hatten Glück bei der Wahl unserer Methode sowie in vielerlei anderer Hinsicht. Ein wiederkehrendes Thema dieses Buches ist, dass Glück in jeder Erfolgsgeschichte eine große Rolle spielt; wenn die Geschichte einen nur etwas anderen Verlauf genommen hätte, dann wäre statt einer bemerkenswerten eine mittelmäßige Leistung herausgekommen. Unsere Geschichte bildete da keine Ausnahme.

Die Reaktion auf unsere Arbeit fiel nicht einhellig positiv aus. Insbesondere wurde an unserer Fokussierung auf kognitive Verzerrungen

kritisiert, dass sie eine unangemessen negative Sichtweise des menschlichen Denkens nahelege.[3] Wie in der Wissenschaft üblich, haben einige Forscher unsere Ideen weiterentwickelt, und andere legten plausible Alternativen vor.[4] Im Großen und Ganzen aber ist die Idee, dass unser Denken anfällig ist für systematische Fehler, heute allgemein anerkannt. Unsere Forschungsarbeiten über Urteilsprozesse hatten einen viel stärkeren Einfluss auf die Sozialwissenschaften, als wir es ursprünglich für möglich hielten.

Unmittelbar nachdem wir unsere neue Theorie über Urteilsprozesse vollständig ausgearbeitet hatten, wandten wir unsere Aufmerksamkeit der Entscheidungsfindung unter Ungewissheit zu. Wir wollten eine psychologische Theorie über die Entscheidungsfindung bei einfachen Glücksspielen erstellen. Zum Beispiel: Würden Sie eine Wette auf einen Münzwurf eingehen, bei der Sie 130 Dollar gewinnen, wenn die Münze mit Kopf nach oben liegt, und 100 Dollar verlieren, wenn die Zahl zu sehen ist? Diese elementaren Entscheidungen wurden schon seit Langem dazu benutzt, allgemeine Fragen der Entscheidungsfindung zu untersuchen, wie zum Beispiel das relative Gewicht, das Menschen sicheren Sachverhalten und ungewissen Ergebnissen zumessen. Unsere Methode änderte sich nicht: Wir verbrachten viele Tage damit, uns Entscheidungsprobleme auszudenken und zu untersuchen, ob unsere intuitiven Präferenzen der Logik der Entscheidung entsprachen. Wie bei Urteilsprozessen beobachteten wir auch in unseren eigenen Entscheidungen systematische Verzerrungen – intuitive Präferenzen, die ständig gegen die Regeln rationaler Entscheidungsfindung verstießen. Fünf Jahre nach dem *Science*-Artikel veröffentlichten wir den Beitrag »Prospect Theory: An Analysis of Decision Under Risk« (»Neue Erwartungstheorie: Eine Analyse der Entscheidungsfindung unter Unsicherheit«), eine Theorie der Entscheidungsfindung, die aus einigen Gründen einflussreicher ist als unsere Arbeit über Urteilsprozesse und eine der Grundlagen der Verhaltensökonomik darstellt.

Bis die geografische Trennung unsere weitere Zusammenarbeit vereitelte, genossen Amos und ich das außerordentliche Glück, dass wir unsere intellektuellen Fähigkeiten kombinieren konnten und so zu Leistungen in der Lage waren, die jeder von uns für sich allein nicht hätte erbringen können, und dass wir eine Beziehung hatten, die unsere Arbeit kurzweilig und produktiv machte. Unsere Zusammenarbeit auf dem Gebiet der Urteils-

und Entscheidungstheorie war der Grund dafür, dass mir 2002 der Nobelpreis zuerkannt wurde, den sich Amos mit mir geteilt hätte, wäre er nicht 1996 im Alter von 59 Jahren gestorben.[5]

## Wo wir heute stehen

In diesem Buch will ich nicht die Ergebnisse der frühen Forschungsarbeiten, die Amos und ich gemeinsam durchführten, darlegen, denn dies haben viele Autoren im Lauf der Jahre bereits mit großer Sachkunde getan. Mir geht es hier vor allem darum, auf der Grundlage der neuesten Entwicklungen in der kognitiven und der Sozialpsychologie einen Überblick über die Funktionsmechanismen des menschlichen Denkens zu geben. Zu den wichtigsten Entwicklungen gehört, dass wir heute die Stärken und die Schwächen des intuitiven Denkens verstehen.

Abgesehen von der beiläufigen Feststellung, dass Urteilsheuristiken »recht nützlich sind, aber manchmal zu schwerwiegenden, systematischen Fehlern führen«, haben Amos und ich uns nicht mit richtigen Intuitionen beschäftigt. Wir konzentrierten uns auf kognitive Verzerrungen, weil wir sie an sich interessant fanden, aber auch, weil sie Beweise für die Urteilsheuristiken lieferten. Wir fragten uns nicht, ob alle intuitiven Urteile unter Ungewissheit von den von uns erforschten Heuristiken hervorgebracht werden; mittlerweile wissen wir, dass dies nicht der Fall ist. Insbesondere die richtigen Intuitionen von Experten lassen sich besser mit den Folgen langjähriger Übung als mit Heuristiken erklären. [6] Wir können heute ein differenzierteres und ausgewogeneres Bild zeichnen, in dem Kompetenzen und Heuristiken alternative Quellen intuitiver Urteile und Entscheidungen sind.

Der Psychologe Gary Klein erzählt die Geschichte eines Teams von Feuerwehrmännern, die in ein Haus eindrangen, in dem die Küche in Flammen stand.[7] Kaum dass sie mit den Löscharbeiten begonnen hatten, hörte sich ihr Anführer selbst »Raus hier!« schreien, ohne zu wissen, warum. Die Decke stürzte ein, just als der letzte Feuerwehrmann den Raum verlassen hatte. Erst im Anschluss wurde dem Zugführer bewusst, dass das Feuer ungewöhnlich leise und seine Ohren ungewöhnlich heiß gewesen waren. Zusammen lösten diese Eindrücke das aus, was er einen »sechsten Sinn für Gefahren« nannte. Er hatte keine Ahnung, was nicht stimmte, aber er wusste, dass etwas nicht stimmte. Es zeigte sich, dass der Brandherd

nicht in der Küche gewesen war, sondern im Kellergeschoss, unterhalb der Stelle, wo die Männer gestanden hatten.

Wir alle haben schon solche Geschichten über das sichere Gespür von Experten gehört: Der Schachgroßmeister, der an einer Schachpartie unter freiem Himmel vorbeigeht und, ohne stehen zu bleiben, verkündet: »Weiß setzt in drei Zügen matt.« Oder der Arzt, der nach einem einzigen Blick auf einen Patienten eine komplexe Diagnose stellt. Die Intuition von Experten kommt uns wie ein Wunder vor, aber das ist nicht der Fall. Tatsächlich vollbringt jeder von uns viele Male pro Tag intuitive Meisterleistungen. Die meisten von uns können mit absoluter Sicherheit schon im ersten Wort eines Anrufers seine Wut spüren, beim Betreten eines Raumes erkennen, dass sie der Gegenstand der Unterhaltung sind, und rasch auf subtile Anzeichen dafür reagieren, dass der Fahrer des Autos auf der Nebenspur gefährlich ist. Unsere alltäglichen intuitiven Fähigkeiten sind nicht weniger wunderbar als die verblüffenden Ahnungen eines erfahrenen Feuerwehrmannes oder Arztes – nur weiter verbreitet.

Die Psychologie richtiger Intuitionen hat nichts Magisches an sich. Die vielleicht beste prägnante Beschreibung stammt von dem großen Herbert Simon, der Schachgroßmeister psychologischen Tests unterzog und feststellte, dass sie nach vielen Tausend Stunden Übung die Figuren auf dem Brett anders sehen als wir.[8] Man spürt deutlich Simons Unverständnis für die Mythologisierung der Experten-Intuition, wenn er schreibt: »Die Situation liefert einen Hinweisreiz; dieser Hinweisreiz gibt dem Experten Zugang zu Informationen, die im Gedächtnis gespeichert sind, und diese Informationen geben ihm die Antwort. Intuition ist nicht mehr und nicht weniger als Wiedererkennen.«[9]

Wir sind nicht überrascht, wenn ein zweijähriges Kleinkind einen Hund betrachtet und »Wauwau!« sagt, weil wir an das Wunder gewöhnt sind, dass Kinder lernen, Gegenstände wiederzuerkennen und zu benennen. Simon will darauf hinaus, dass die intuitiven Meisterleistungen von Experten den gleichen Charakter haben. Zu richtigen Intuitionen kommt es dann, wenn Experten gelernt haben, vertraute Elemente in einer neuen Situation wiederzuerkennen und in einer Weise zu handeln, die ihr angemessen ist. Gute intuitive Urteile tauchen mit der gleichen Unmittelbarkeit im Bewusstsein auf wie »Wauwau!«.

Leider entspringen nicht alle Intuitionen von Fachleuten echtem Sachverstand. Vor vielen Jahren besuchte ich den Leiter der Vermögensverwaltung eines großen Finanzdienstleisters, der mir sagte, er habe gerade einige zehn Millionen Dollar in Aktien der Ford Motor Company investiert. Als ich ihn fragte, wie er zu diesem Entschluss gelangt sei, antwortete er, er sei kürzlich auf einer Automesse gewesen und das, was er gesehen habe, habe ihn beeindruckt. »Mann, die wissen, wie man ein Auto baut!«, war seine Erklärung. Er ließ keinen Zweifel daran, dass er seinem Bauchgefühl vertraute, und war zufrieden mit sich und mit seiner Entscheidung. Ich fand es bemerkenswert, dass er anscheinend die eine Frage, die ein Ökonom als relevant erachten würde, nicht in Betracht gezogen hatte: Sind Ford-Aktien gegenwärtig unterbewertet? Stattdessen hörte er auf seine Intuition; er mochte Autos, er mochte das Unternehmen, und er mochte die Vorstellung, Ford-Aktien zu besitzen. Nach allem, was wir über die Treffgenauigkeit beim Stock-Picking – der gezielten Suche nach unterbewerteten Einzeltiteln – wissen, kann man vernünftigerweise davon ausgehen, dass er nicht wusste, was er tat.

Die spezifischen Heuristiken, die Amos und ich erforschten, helfen uns kaum, zu verstehen, was den Manager dazu veranlasste, in Ford-Aktien zu investieren, aber heute verfügen wir über eine umfassendere Konzeption von Heuristiken, die uns eine gute Erklärung liefert. Ein wichtiger Fortschritt besteht darin, dass Emotionen heute beim Verständnis intuitiver Urteile und Entscheidungen eine viel größere Rolle spielen als in der Vergangenheit. Die Entscheidung des Managers würde heute als Beispiel der Affektheuristik beschrieben werden; dabei handelt es sich um Urteile und Entscheidungen, die unmittelbar und ohne gründliche Überlegung von Gefühlen der Vorliebe und Abneigung bestimmt werden.[10] Wenn die Maschinerie des intuitiven Denkens mit einem Problem konfrontiert ist – der Auswahl eines Schachzugs oder der Entscheidung, in eine bestimmte Aktie zu investieren –, tut sie ihr Bestes. Wenn die Person über sachdienliches Fachwissen verfügt, erkennt sie die Situation, und die intuitive Lösung, die ihr einfällt, wird wahrscheinlich richtig sein. Dies geschieht, wenn ein Schachgroßmeister eine komplexe Position betrachtet: Die wenigen Züge, die ihm sofort einfallen, sind durchweg starke Züge. Wenn die Frage schwierig und eine sachkundige Lösung nicht verfügbar ist, hat die Intuition trotzdem eine Chance: Dem Betreffenden mag schnell eine Antwort einfallen – aber es ist keine Antwort auf die ursprüngliche Frage.

Die Frage, vor der der Manager stand (Soll ich in Ford-Aktien investieren?), war schwierig, aber die Antwort auf eine leichtere und damit zusammenhängende Frage (Mag ich Autos von Ford?) fiel ihm spontan ein und bestimmte seine Entscheidung. Das ist das Wesen intuitiver Heuristiken: Wenn wir mit einer schwierigen Frage konfrontiert sind, beantworten wir stattdessen oftmals eine leichtere, ohne dass wir die Ersetzung bemerken.[11]

Die spontane Suche nach einer intuitiven Lösung scheitert manchmal – es fällt einem weder eine Expertenlösung noch eine heuristische Antwort ein. In solchen Fällen wechseln wir oftmals zu einer langsameren, wohlüberlegten und anstrengenden Form des Denkens. Dies ist das »langsame Denken«, das im Titel erwähnt wird. Schnelles Denken umfasst beide Varianten des intuitiven Denkens – das sachkundige und das heuristische Denken – sowie die vollkommen automatisierten mentalen Wahrnehmungs- und Gedächtnisprozesse, also jene Operationen, die uns in die Lage versetzen, das Objekt auf unserem Schreibtisch als Lampe zu erkennen oder den Namen der russischen Hauptstadt aus dem Gedächtnis abzurufen.

Zahlreiche Psychologen haben in den letzten 25 Jahren die Unterschiede zwischen schnellem und langsamem Denken erforscht. Aus Gründen, die ich im nächsten Kapitel ausführlicher darlege, beschreibe ich mentale Prozesse mit der Metapher von zwei Agenten, System 1 und System 2 genannt, die jeweils schnelles und langsames Denken erzeugen. Ich spreche von den Merkmalen intuitiven und bewussten Denkens, als handele es sich um Merkmale und Dispositionen von zwei Figuren in unserem Geist. Die jüngsten Forschungen deuten darauf hin, dass das intuitive System 1 einflussreicher ist, als dies nach unserem subjektiven Erleben der Fall zu sein scheint, und es ist der geheime Urheber vieler Entscheidungen und Urteile, die wir treffen. Den Schwerpunkt dieses Buches bilden die Funktionsmechanismen von System 1 und die wechselseitigen Einflüsse zwischen System 1 und System 2.

## Was als Nächstes kommt

Dieses Buch besteht aus fünf Teilen. Teil I stellt die Grundelemente einer Urteils- und Entscheidungstheorie auf der Basis zweier Systeme vor. Der Unterschied zwischen den unwillkürlichen Operationen von System 1 und den bewusst gesteuerten Operationen von System 2 wird ausführlich dargelegt, und es wird gezeigt, wie das assoziative Gedächtnis, der Kern von System 1, fortwährend eine kohärente Interpretation dessen konstruiert, was zu jedem beliebigen Zeitpunkt in unserer Welt geschieht. Ich versuche, dem Leser eine Vorstellung von der Komplexität und Reichhaltigkeit der automatischen und oftmals unbewussten Prozesse zu vermitteln, die dem intuitiven Denken zugrunde liegen, und darzustellen, wie man die Urteilsheuristiken auf der Grundlage dieser automatischen Prozesse erklären kann. Ein Ziel ist es, eine Sprache einzuführen, die es erlaubt, präziser über kognitive Prozesse nachzudenken und zu sprechen.

Teil II stellt den aktuellen Wissensstand bei Urteilsheuristiken dar und geht einer wichtigen Frage nach: Weshalb ist es für uns so schwierig, statistisch zu denken? Es fällt uns leicht, assoziativ oder metaphorisch zu denken, wir denken kausal, aber eine statistische Betrachtungsweise erfordert, dass wir an viele Dinge gleichzeitig denken, worauf System 1 nicht ausgelegt ist.

Die Schwierigkeiten, die uns statistisches Denken bereitet, leiten über zu dem Hauptthema von Teil III, der eine rätselhafte Beschränkung unseres Denkens beschreibt: unser übermäßiges Vertrauen in das, was wir zu wissen glauben, und unsere scheinbare Unfähigkeit, das ganze Ausmaß unseres Unwissens und der Unbestimmtheit der Welt zuzugeben. Wir überschätzen tendenziell unser Wissen über die Welt, und wir unterschätzen die Rolle, die der Zufall bei Ereignissen spielt. Überzogenes Vertrauen in die Vorhersagbarkeit der Welt wird durch die illusorische Gewissheit retrospektiver Einsichten gestützt. Nassim Taleb, der Autor des Buches *Der Schwarze Schwan*, hat meine Ansichten zu diesem Thema maßgeblich beeinflusst. Ich hoffe auf Gespräche am Kaffeeautomaten, die die Lektionen, die wir aus der Vergangenheit lernen können, auf intelligente Weise erkunden, während sie der Verlockung rückblickender Verzerrung und der Illusion der Gewissheit widerstehen.

Der Schwerpunkt von Teil IV ist ein Gespräch mit den Wirtschaftswissenschaften über die Natur der Entscheidungsfindung und die Annahme, dass ökonomische Agenten rational handeln. In diesem Abschnitt des Buches stelle ich eine aktualisierte Version der Schlüsselkonzepte der Neuen Erwartungstheorie *(Prospect Theory)* auf der Basis des Zwei-System-Modells vor, jener Entscheidungstheorie, die Amos und ich 1979 publizierten. Daran anschließende Kapitel befassen sich mit verschiedenen Weisen, wie menschliche Entscheidungen von den Regeln der Rationalität abweichen können. Ich befasse mich mit der bedauerlichen Neigung, Probleme isoliert zu behandeln, und mit sogenannten Framing-Effekten (»Einrahmungseffekten«), bei denen Entscheidungen maßgeblich von unwesentlichen Merkmalen der Entscheidungsprobleme beeinflusst werden. Diese Beobachtungen, die sich ohne Weiteres mit den Eigenschaften von System 1 erklären lassen, stellen eine große Herausforderung für die Rationalitätsannahme dar, wie sie in der herrschenden volkswirtschaftlichen Lehre vertreten wird.

Teil V beschreibt neueste Forschungen, die eine Unterscheidung zwischen zwei Formen des Selbst eingeführt haben – dem erlebenden Selbst und dem sich erinnernden Selbst –, die nicht die gleichen Interessen verfolgen. So können wir beispielsweise Menschen zwei schmerzhaften Erfahrungen aussetzen. Eine dieser Erfahrungen ist deutlich unangenehmer als die andere, weil sie länger dauert. Aber die automatische Bildung von Erinnerungen – ein Merkmal von System 1 – folgt bestimmten Regeln, die wir uns zunutze machen können, sodass uns die unangenehmere Episode in besserer Erinnerung bleibt. Wenn Menschen vor der Entscheidung stehen, welche Episode sie wiederholen sollen, werden sie unwillkürlich von ihrem erinnernden Selbst geleitet und setzen sich (ihr erlebendes Selbst) unnötigen Schmerzen aus. Die Differenzierung zwischen zwei Formen des Selbst wird auf die Messung des Wohlbefindens angewandt, wo wir erneut feststellen, dass das, was das erlebende Selbst glücklich macht, nicht das Gleiche ist, was das erinnernde Selbst befriedigt. Die Frage, wie beide Selbste in einem Körper Befriedigung finden können, wirft eine Reihe von Schwierigkeiten auf, sowohl für Individuen als auch für Gesellschaften, die in der Verwirklichung des Gemeinwohls eine politische Zielsetzung sehen.

In einem abschließenden Kapitel betrachten wir, in umgekehrter Reihenfolge, die Tragweite der drei in diesem Buch vorgenommenen Unterscheidungen: zwischen dem erlebenden und dem erinnernden Selbst,

zwischen den Konzeptionen des Agenten in der klassischen Nationalökonomie und in der Verhaltensökonomik (die Anleihen bei der Psychologie macht) sowie zwischen dem unwillkürlichen System 1 und dem mühevollen System 2. Ich kehre zurück zu den Vorteilen eines »aufgeklärten« Bürotratschs und zu der Frage, was Organisationen tun können, um die Qualität von Urteilen und Entscheidungen, die Bedienstete in ihrem Auftrag treffen, zu verbessern. Zwei Aufsätze, die ich zusammen mit Amos geschrieben habe, sind im Anhang abgedruckt. Der zweite Aufsatz, der 1984 veröffentlicht wurde, resümiert die Neue Erwartungstheorie und unsere Studien über Framing-Effekte. Diese Artikel stellen jene wissenschaftlichen Beiträge dar, auf die das Nobelpreiskomitee Bezug genommen hat – und es wird Sie vielleicht überraschen, wie einfach sie sind. Ihre Lektüre wird Ihnen eine Vorstellung davon vermitteln, wie viel wir schon seit langer Zeit wissen, aber auch davon, was wir in den letzten Jahrzehnten dazugelernt haben.

# TEIL I

## Zwei Systeme

# *1. Die Figuren der Geschichte*

Wenn Sie dieses Bild betrachten, können Sie Ihr Denken im automatischen Modus beobachten.

Abbildung 1

Ihr Erleben beim Betrachten des Gesichts der Frau verknüpft nahtlos das, was wir normalerweise Sehen und intuitives Denken nennen. So sicher und so schnell, wie Sie sehen, dass die Frau dunkles Haar hat, erkennen Sie, dass sie wütend ist. Außerdem weist das, was Sie sehen, in die Zukunft. Sie spüren, dass diese Frau kurz davor ist, einige sehr unfreundliche Worte zu äußern, vermutlich mit lauter, schriller Stimme. Eine Vorahnung dessen, was sie als Nächstes tun wird, stellt sich automatisch und mühelos in Ihrem Bewusstsein ein. Sie hatten nicht die Absicht, ihren Gemütszustand einzuschätzen oder das, was sie als Nächstes tun würde, zu antizipieren, und Ihre Reaktion auf das Bild fühlte sich für Sie nicht wie etwas an, das Sie bewusst taten. Es widerfuhr Ihnen einfach. Es war ein Fall von schnellem Denken.

Betrachten Sie jetzt das folgende Problem:

17 × 24

Ihnen ist sofort klar, dass dies eine Multiplikationsaufgabe ist, und vermutlich wissen Sie, dass Sie sie mit Bleistift und Papier – wenn nicht sogar ohne – lösen können. Sie besitzen auch ein vages intuitives Wissen über die Spannweite möglicher Ergebnisse. Sie erkennen rasch, dass sowohl 12 609 als auch 123 unplausibel sind. Doch ohne eine gewisse Zeit auf das Problem zu verwenden, wären Sie nicht sicher, dass die Antwort nicht 568 ist. Eine exakte Lösung fällt Ihnen nicht sofort ein, und Sie haben den Eindruck, dass Sie entscheiden können, ob Sie die Berechnung durchführen wollen oder nicht. Wenn Sie es noch nicht getan haben, sollten Sie das Multiplikationsproblem jetzt zu lösen versuchen und wenigstens einen Teil davon abschließen.

Während Sie eine Reihe von Rechenschritten absolvieren, erleben Sie langsames Denken. Als Erstes haben Sie das kognitive Multiplikationsprogramm, das Sie in der Schule lernten, aus Ihrem Gedächtnis abgerufen – anschließend haben Sie dieses umgesetzt. Die Berechnung ist mühsam. Sie spüren, wie belastend es ist, viel Stoff im Gedächtnis zu behalten, weil Sie nicht den Überblick darüber verlieren dürfen, wo Sie gerade waren und wohin Sie wollen, während Sie das Zwischenergebnis im Geist festhalten. Dieser Prozess ist geistige Arbeit: er erfordert zielgerichtete Anstrengung und Strukturierung, und er ist daher ein Prototyp langsamen Denkens. Die Berechnung ist nicht nur ein mentaler Vorgang; auch Ihr Körper ist daran beteiligt. Ihre Muskeln spannen sich an, Ihr Blutdruck steigt und Ihr Herzschlag ebenfalls. Jemand, der Ihre Augen genau beobachten würde, während Sie mit dem Problem beschäftigt sind, würde sehen, wie sich Ihre Pupillen weiten. Sobald Sie mit der Arbeit fertig sind – wenn Sie die Antwort gefunden haben (die übrigens 408 lautet) oder wenn Sie aufgegeben haben –, schrumpfen Ihre Pupillen wieder auf normale Größe.

## Zwei Systeme

Seit mehreren Jahrzehnten erforschen Psychologen intensiv die beiden Denkmodi, für die das Bild der zornigen Frau einerseits und das Multiplikationsproblem andererseits stehen, und sie haben zahlreiche Bezeichnungen dafür vorgeschlagen. [1] Ich verwende die Termini, die ursprünglich von den Psychologen Keith Stanovich und Richard West eingeführt wurden, und ich werde entsprechend zwei kognitive Systeme unterscheiden, System 1 und System 2.

- *System 1* arbeitet automatisch und schnell, weitgehend mühelos und ohne willentliche Steuerung.
- *System 2* lenkt die Aufmerksamkeit auf die anstrengenden mentalen Aktivitäten, die auf sie angewiesen sind, darunter auch komplexe Berechnungen. Die Operationen von System 2 gehen oftmals mit dem subjektiven Erleben von Handlungsmacht, Entscheidungsfreiheit und Konzentration einher.[2]

Die Bezeichnungen System 1 und System 2 sind in der Psychologie allgemein geläufig, aber ich gehe in diesem Buch, das Sie als ein Psychodrama mit zwei Figuren lesen können, weiter als die meisten.

Wenn wir an uns selbst denken, identifizieren wir uns mit System 2, dem bewussten, logisch denkenden Selbst, das Überzeugungen hat, Entscheidungen trifft und sein Denken und Handeln bewusst kontrolliert. Obwohl System 2 von sich selbst glaubt, im Zentrum des Geschehens zu stehen, ist das unwillkürliche System 1 der Held dieses Buches. In System 1 entstehen spontan die Eindrücke und Gefühle, die die Hauptquellen der expliziten Überzeugungen und bewussten Entscheidungen von System 2 sind. Die automatischen Operationen von System 1 erzeugen erstaunlich komplexe Muster von Vorstellungen, aber nur das langsamere System 2 kann in einer geordneten Folge von Schritten Gedanken konstruieren. Ich beschreibe auch Umstände, unter denen System 2 die Kontrolle übernimmt, indem es die ungezügelten Impulse und Assoziationen von System 1 verwirft. Man kann die beiden Systeme mit Akteuren vergleichen, die jeweils individuelle Fähigkeiten, Beschränkungen und Funktionen

aufweisen. In näherungsweiser Reihenfolge der Komplexität sind hier einige Beispiele für die automatischen Aktivitäten aufgelistet, die System 1 zugeschrieben werden:

– Erkennen Sie, dass ein Gegenstand weiter entfernt ist als ein anderer.
– Wenden Sie sich der Quelle eines plötzlichen Geräuschs zu.
– Vervollständigen Sie den Ausdruck »Brot und …«
– Ziehen Sie ein »angewidertes Gesicht«, wenn man Ihnen ein grauenvolles Bild zeigt.
– Hören Sie die Feindseligkeit aus einer Stimme heraus.
– Beantworten Sie: 2 + 2 = ?
– Lesen Sie Wörter auf großen Reklameflächen.
– Fahren Sie mit einem Auto über eine leere Straße.
– Finden Sie einen starken Schachzug (wenn Sie Schachmeister sind).
– Verstehen Sie einfache Sätze.
– Erkennen Sie, dass eine »sanftmütige und ordentliche Person mit großer Liebe zum Detail« einem beruflichen Stereotyp entspricht.

All diese mentalen Ereignisse gehören zur gleichen Kategorie wie die wütende Frau – sie geschehen automatisch und weitgehend mühelos. Zu den Funktionen von System 1 gehören angeborene Fähigkeiten, die wir mit anderen Tieren gemeinsam haben. Wir werden mit der Fähigkeit geboren, unsere Umwelt wahrzunehmen, Gegenstände zu erkennen, unsere Aufmerksamkeit zu steuern, Verluste zu vermeiden und uns vor Spinnen zu fürchten. Andere mentale Aktivitäten werden durch lange Übung zu schnellen, automatisierten Routinen. System 1 hat Assoziationen zwischen Vorstellungen gelernt (die Hauptstadt Frankreichs?); es hat auch Fähigkeiten gelernt, wie etwa das Lesen und Verstehen von Nuancen sozialer Situationen. Einige Fähigkeiten, wie etwa das Finden starker Schachzüge, werden nur von spezialisierten Experten erworben. Andere sind weitverbreitet. Um die Ähnlichkeit eines Persönlichkeitsprofils mit einem beruflichen Stereotyp zu erkennen, benötigt man ein umfassendes Wissen um Sprache und Kultur, das die meisten von uns besitzen. Das Wissen ist im Gedächtnis gespeichert und wird ohne Intention und ohne Anstrengung abgerufen.

Etliche der mentalen Aktivitäten in der Liste erfolgen vollkommen unwillkürlich. Man kann sich nicht davon abhalten, einfache Sätze in seiner Muttersprache zu verstehen oder sich zu einem unerwarteten lauten

Geräusch umzudrehen, noch kann man sich daran hindern, zu wissen, dass 2 + 2 = 4 ist, oder an Paris zu denken, wenn von der Hauptstadt Frankreichs die Rede ist. Andere Aktivitäten, wie das Kauen, sind der willentlichen Kontrolle zugänglich, werden für gewöhnlich aber von einem Autopiloten gesteuert. Beide Systeme sind an der Aufmerksamkeitssteuerung beteiligt. Die Hinwendung zu einem lauten Geräusch ist normalerweise eine unwillkürliche Operation von System 1, das sofort die willkürliche Aufmerksamkeit von System 2 mobilisiert. Vielleicht können Sie dem Impuls widerstehen, sich der Quelle einer lauten und unverschämten Bemerkung bei einer übervollen Party zuzuwenden, aber selbst wenn sich Ihr Kopf nicht bewegt, ist Ihre Aufmerksamkeit zunächst auf die Quelle gerichtet, zumindest eine Zeit lang. Allerdings kann die Aufmerksamkeit von einem ungewollten Fokus abgezogen werden, hauptsächlich dadurch, dass man sich gezielt auf etwas anderes konzentriert. Die höchst vielfältigen Aktivitäten von System 2 haben ein Merkmal gemeinsam: Sie erfordern Aufmerksamkeit, und sie werden gestört, wenn die Aufmerksamkeit abgezogen wird.

Hier sind einige Beispiele:

– Sich bei einem Wettlauf auf den Startschuss einstellen.
– Die Aufmerksamkeit auf die Clowns in einem Zirkus richten.
– Sich auf die Stimme einer bestimmten Person in einem überfüllten und sehr lauten Raum konzentrieren.
– Nach einer Frau mit weißem Haar Ausschau halten.
– Das Gedächtnis durchsuchen, um ein ungewohntes Geräusch zu identifizieren.
– Schneller gehen, als Sie es normalerweise tun.
– Die Angemessenheit Ihres Verhaltens in einer sozialen Situation überwachen.
– Zählen, wie oft der Buchstabe a auf einer Textseite vorkommt.
– Jemandem seine Telefonnummer mitteilen.
– In eine schmale Lücke einparken (für die meisten Leute, bis auf die Mitarbeiter von Kfz-Werkstätten).
– Zwei Waschmaschinen auf das bessere Preis-Leistungs-Verhältnis hin vergleichen.
– Eine Steuererklärung anfertigen.
– Die Gültigkeit einer komplexen logischen Beweisführung überprüfen.

In all diesen Situationen muss man sich konzentrieren, und die Leistung, die man erbringt, fällt schlechter aus, wenn man nicht dazu bereit ist oder wenn die Aufmerksamkeit in unangemessener Weise fokussiert wird. System 2 besitzt die Fähigkeit, die Funktionsweise von System 1 in gewissem Umfang zu verändern, indem es die normalerweise automatischen Funktionen von Aufmerksamkeit und Gedächtnis programmiert. Wenn man zum Beispiel auf einem belebten Bahnhof auf einen Verwandten wartet, kann man sich willentlich darauf konzentrieren, nach einer weißhaarigen Frau oder einem bärtigen Mann Ausschau zu halten, und dadurch die Wahrscheinlichkeit erhöhen, dass man den Verwandten schon von fern bemerkt. Man kann in seinem Gedächtnis gezielt nach Hauptstädten, die mit dem Buchstaben N beginnen, oder nach den Titeln existenzialistischer französischer Romane suchen. Und wenn man ein Auto am Londoner Heathrow Airport mietet, wird einen der Mitarbeiter vermutlich daran erinnern, dass »wir hier links fahren«. In all diesen Fällen wird man aufgefordert, etwas zu tun, was nicht spontan geschieht, und man stellt fest, dass die Aufrechterhaltung einer bestimmten Einstellung zumindest ein gewisses Maß an andauernder Anstrengung erfordert. Der im Englischen häufig verwendete Ausdruck *to pay attention* (wörtlich: »Aufmerksamkeit zahlen«, also schenken) ist passend: Man verfügt über ein begrenztes Aufmerksamkeitsbudget, das man auf verschiedene Aktivitäten verteilen kann; wenn man versucht, sein Budget zu überschreiten, misslingt dies. Es ist das Kennzeichen anstrengender Aktivitäten, dass sie einander überlagern, und aus diesem Grund ist es schwer oder unmöglich, mehrere gleichzeitig auszuführen. Wir können nicht das Produkt von 17 × 24 berechnen, während wir bei dichtem Verkehr links abbiegen, und man sollte es mit Sicherheit gar nicht erst versuchen. Man kann mehrere Dinge gleichzeitig tun, aber nur wenn sie einfach und anspruchslos sind. Es ist wahrscheinlich ungefährlich, wenn Sie mit Ihrem Beifahrer plaudern, während Sie auf einer leeren Fernstraße fahren, und viele Eltern haben entdeckt – vielleicht mit einem leicht schlechten Gewissen –, dass sie einem Kind eine Geschichte vorlesen können, während sie an etwas anderes denken.

Jeder Mensch ist sich seiner begrenzten Aufmerksamkeitskapazität irgendwie bewusst, und unser soziales Verhalten berücksichtigt diese Beschränkungen. Wenn der Fahrer eines Autos auf einer schmalen Straße einen Laster überholt, verstummen die erwachsenen Mitfahrer aus

nachvollziehbaren Gründen jäh. Sie wissen, dass es keine gute Idee ist, den Fahrer abzulenken, und sie nehmen auch an, dass er vorübergehend taub ist und nicht hören würde, was sie sagen.

Die intensive Konzentration auf eine Aufgabe kann Menschen tatsächlich blind für Stimuli machen, die normalerweise ihre Aufmerksamkeit erregen würden. Die spektakulärste Demonstration dafür lieferten Christopher Chabris und Daniel Simons in ihrem Buch *Der unsichtbare Gorilla*. Sie produzierten einen kurzen Film über zwei Mannschaften, die sich Basketbälle zuspielten, wobei ein Team weiße Hemden trug und das andere schwarze. Die Betrachter des Films werden aufgefordert, die Zahl der Ballwechsel der weißen Mannschaft zu zählen und die schwarzen Spieler zu ignorieren. Das ist eine schwierige Aufgabe, die volle Konzentration verlangt. Ungefähr in der Mitte des Videos taucht eine Frau auf, die als Gorilla verkleidet ist, überquert das Spielfeld und verschwindet wieder. Der Gorilla ist neun Sekunden lang zu sehen. Tausende von Menschen haben sich das Video angeschaut, und etwa der Hälfte von ihnen fällt nichts Ungewöhnliches auf. Ursache dieser Blindheit ist die Zählaufgabe – und insbesondere die Anweisung, eines der Teams zu ignorieren. Niemand, der das Video ohne diese Aufgabe betrachtet, würde den Gorilla übersehen. Sehen und Sich-orientieren sind automatische Funktionen von System 1, aber sie sind darauf angewiesen, dass dem relevanten Stimulus eine gewisse Aufmerksamkeit zugewendet wird. Die Autoren weisen darauf hin, dass die bemerkenswerteste Beobachtung ihrer Studie darin besteht, dass Menschen deren Ergebnisse sehr überraschend finden. Tatsächlich sind die Filmbetrachter, die den Gorilla nicht gesehen haben, zunächst fest davon überzeugt, dass er nicht da war – sie können sich nicht vorstellen, dass ihnen ein so auffallendes Ereignis entgangen ist. Die Gorilla-Studie verdeutlicht zwei wichtigen Tatsachen über mentale Prozesse: Wir können gegenüber dem Offensichtlichen blind sein, und wir sind darüber hinaus blind für unsere Blindheit.

## Der Gang der Handlung – ein kurzer Überblick

Die Wechselwirkung zwischen den beiden Systemen ist ein ständig wiederkehrendes Thema dieses Buches, und es dürfte hilfreich sein, den Handlungsverlauf kurz zusammenzufassen. In der Geschichte, die ich erzählen werde, sind sowohl System 1 als auch System 2 immer aktiv, wenn wir wach sind. System 1 läuft automatisch, und System 2 befindet sich normalerweise in einem angenehmen Modus geringer Anstrengung, in dem nur ein Teil seiner Kapazität in Anspruch genommen wird. System 1 generiert fortwährend Vorschläge für System 2: Eindrücke, Intuitionen, Absichten und Gefühle. Wenn Eindrücke und Intuitionen von System 2 unterstützt werden, werden sie zu Überzeugungen, und Impulse werden zu willentlich gesteuerten Handlungen. Wenn alles glattläuft, was meistens der Fall ist, macht sich System 2 die Vorschläge von System 1 ohne größere Modifikationen zu eigen. Im Allgemeinen vertraut man seinen Eindrücken und gibt seinen Wünschen nach, und das ist in Ordnung so – für gewöhnlich.

Wenn System 1 in Schwierigkeiten gerät, fordert es von System 2 eine detailliertere und spezifischere Verarbeitung an, die das anstehende Problem möglicherweise lösen wird. System 2 wird mobilisiert, wenn eine Frage auftaucht, für die System 1 keine Antwort bereitstellt, wie es vermutlich der Fall war, als Sie mit dem Multiplikationsproblem 17 × 24 konfrontiert waren. Auch wenn etwas Überraschendes geschieht, kommt es zu einem jähen Anstieg der bewussten Aufmerksamkeit. System 2 wird aktiviert, wenn ein Ereignis registriert wird, das gegen das Weltmodell von System 1 verstößt. In dieser Welt gibt es keine hüpfenden Lampen, bellenden Katzen und keine Gorillas, die über Basketballfelder laufen. Das Gorilla-Experiment zeigt, dass der unerwartete Reiz nur dann wahrgenommen wird, wenn eine gewisse Aufmerksamkeit auf ihn gelenkt wird. Die Überraschung aktiviert und orientiert dann die Aufmerksamkeit: Man heftet den Blick auf den Reiz, und man durchsucht sein Gedächtnis nach einer Geschichte, die dem unerwarteten Ereignis einen Sinn gibt. System 2 ist auch für die fortwährende Überwachung des Verhaltens zuständig – die Kontrolle, die gewährleistet, dass man höflich bleibt, wenn man wütend ist, und dass man hellwach bleibt, wenn man nachts Auto fährt. System 2 wird

hochgefahren, sobald es einen drohenden Fehler bemerkt. Erinnern Sie sich an eine Gelegenheit, bei der Ihnen beinahe eine unverschämte Bemerkung über die Lippen gekommen wäre, und erinnern Sie sich, wie hart Sie darum ringen mussten, die Kontrolle wiederzuerlangen. Kurz und gut, der größte Teil dessen, was Sie (Ihr System 2) denken und tun, geht aus System 1 hervor, aber System 2 übernimmt, sobald es schwierig wird, und es hat normalerweise das letzte Wort.

Die Arbeitsteilung zwischen System 1 und System 2 ist höchst effizient: Sie minimiert den Aufwand und optimiert die Leistung. Diese Regelung funktioniert meistens gut, weil System 1 im Allgemeinen höchst zuverlässig arbeitet: seine Modelle vertrauter Situationen sind richtig, seine kurzfristigen Vorhersagen sind in der Regel ebenfalls zutreffend, und seine anfänglichen Reaktionen auf Herausforderungen sind prompt und im Allgemeinen angemessen. Die Leistungsfähigkeit von System 1 wird jedoch durch kognitive Verzerrungen beeinträchtigt, systematische Fehler, für die es unter spezifischen Umständen in hohem Maße anfällig ist. Wie wir sehen werden, beantwortet es manchmal Fragen, die leichter sind als jene, die ihm gestellt wurden, und es versteht kaum etwas von Logik und Statistik. Eine weitere Beschränkung von System 1 besteht darin, dass es nicht abgeschaltet werden kann. Wenn Sie auf einem Bildschirm ein Wort in einer Sprache sehen, die Sie kennen, lesen Sie es – es sei denn, Ihre Aufmerksamkeit wird von etwas anderem vollkommen in Beschlag genommen. [3]

## Konflikt

Abbildung 2 zeigt eine Variante eines klassischen Experiments, das einen Konflikt zwischen den beiden Systemen erzeugt.[4] Sie sollten diese Übung selbst ausprobieren, bevor Sie weiterlesen.

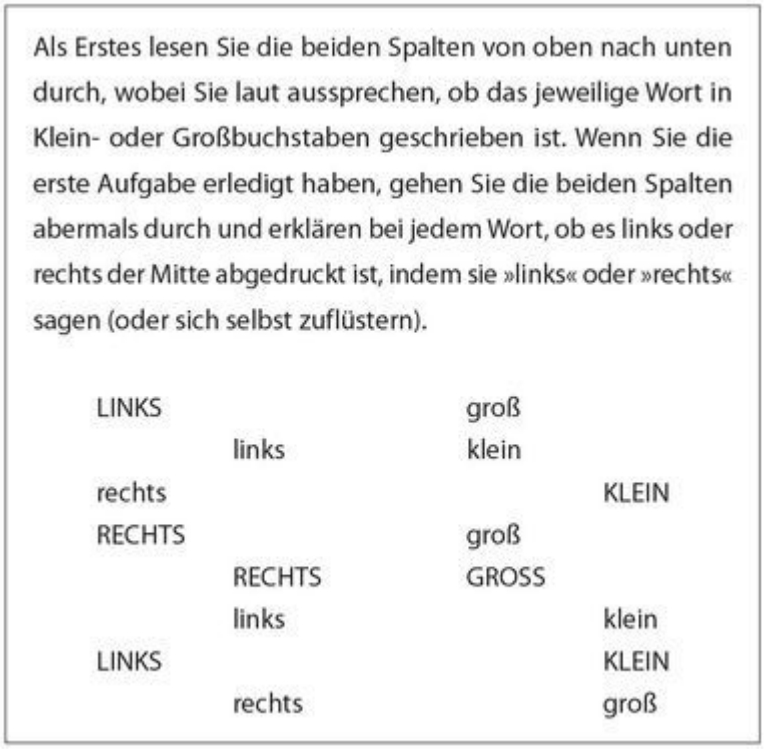


**Abbildung 2**

Wahrscheinlich haben Sie in beiden Aufgaben die richtigen Wörter gesagt, und Sie haben vermutlich bemerkt, dass einige Teile jeder Aufgabe viel leichter als andere waren. Als Sie Groß- und Kleinschreibung benannten, war die linke Spalte leicht, während Sie in der rechten Spalte langsamer wurden und vielleicht ins Stottern und Stocken kamen. Als Sie die Stellung der Wörter angeben sollten, war die linke Spalte schwierig und die rechte Spalte war viel einfacher.

Diese Aufgaben beanspruchen System 2, weil Sie beim Betrachten einer Spalte von Wörtern für gewöhnlich nicht »groß«, »klein« oder »rechts«, »links« sagen. Um sich auf die Aufgabe vorzubereiten, haben Sie unter

anderem Ihr Gedächtnis so programmiert, dass Ihnen die relevanten Wörter (»groß« und »klein« für die erste Aufgabe) »auf der Zunge« lagen. Die Priorisierung der gewählten Wörter ist effektiv, und der schwachen Versuchung, andere Wörter zu lesen, konnten Sie recht leicht widerstehen, als Sie die erste Spalte durchgingen. Aber bei der zweiten Spalte war es anders, weil sie Wörter enthielt, auf die Sie eingestellt waren und die Sie nicht ignorieren konnten. Sie konnten größtenteils richtig antworten, aber es kostete Sie Mühe, die konkurrierende Antwort zu überwinden, und es bremste Sie aus. Sie erlebten einen Konflikt zwischen einer Aufgabe, die Sie erledigen wollten, und einer automatischen Antwort, die damit interferierte.

Konflikte zwischen einer automatischen Reaktion und dem Willen, die Kontrolle zu behalten, kommen in unserem Leben häufig vor. Wir alle haben schon erlebt, wie wir gegen den Impuls ankämpften, das seltsam angezogene Pärchen am Nachbartisch in einem Restaurant anzugaffen. Wir wissen auch alle, wie es ist, wenn man sich dazu zwingen muss, seine Aufmerksamkeit auf ein langweiliges Buch zu richten, und man immer wieder an die Stelle zurückkehrt, wo man den Faden verloren hat. In Gegenden mit harten Wintern erinnern sich viele Fahrer daran, wie ihr Wagen auf dem Eis ins Rutschen kam, sich nicht mehr kontrollieren ließ und sie darum rangen, die gründlich einstudierten Regeln zu befolgen, die das genaue Gegenteil dessen waren, was sie intuitiv tun würden: »In die Richtung lenken, in die der Wagen schleudert, und auf keinen Fall bremsen!« Und jeder Mensch hat schon erlebt, dass er jemandem *nicht* gesagt hat, er solle sich zum Teufel scheren. Eine der Aufgaben von System 2 besteht darin, die Impulse von System 1 zu überwinden. Anders gesagt, System 2 ist für die Selbstbeherrschung zuständig.

## Illusionen

Um die Autonomie von System 1 und den Unterschied zwischen Eindrücken und Überzeugungen zu verstehen, sollten Sie sich Abbildung 3 genau ansehen. Das Bild ist nicht weiter bemerkenswert: zwei horizontale Linien unterschiedlicher Länge, die mit Pfeilspitzen oder »Schwanzflossen« versehen sind, die in unterschiedliche Richtungen zeigen. Die untere Linie ist scheinbar länger als die darüber. Das sehen wir alle, und wir glauben spontan, was wir sehen. Aber wenn Sie dieses Bild schon einmal gesehen haben, wissen Sie, dass es sich um die berühmte Müller-Lyer-Illusion handelt. Wie Sie leicht selbst überprüfen können, indem Sie die horizontalen Linien mit einem Lineal nachmessen, sind beide tatsächlich genau gleich lang.

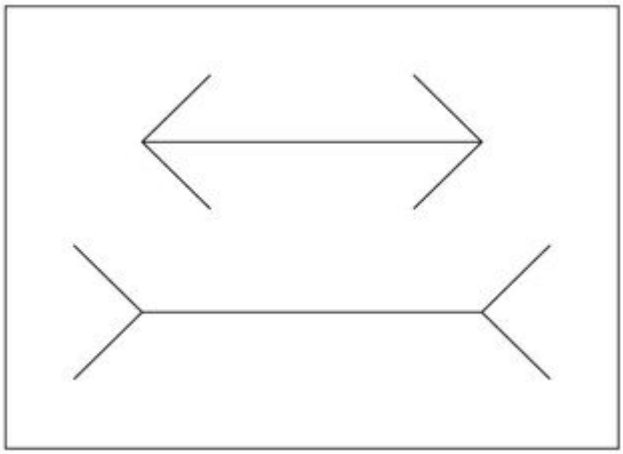

**Abbildung 3**

Jetzt, nachdem Sie die Linien gemessen haben, haben Sie – Ihr System 2, das mit Bewusstsein begabte Wesen, das Sie »ich« nennen – eine neue Überzeugung: Sie *wissen*, dass die Linien gleich lang sind. Wenn Sie nach ihrer Länge gefragt werden, sagen Sie, was Sie wissen. Aber die untere Linie *erscheint* nach wie vor als die längere. Sie haben sich entschieden, der Messung zu glauben, aber Sie können System 1 nicht von seiner gewohnten Aktivität abhalten; Sie können nicht durch einen Willensentschluss Ihre Wahrnehmung so verändern, dass Sie die Linien als gleich lang sehen, obwohl Sie wissen, dass sie gleich lang sind. Um nicht der Illusion zu

erliegen, können Sie nur eines tun: Sie müssen lernen, Ihren Wahrnehmungen der Länge von Linien zu misstrauen, wenn sie mit »Pfeilspitzen« oder »Schwanzflossen« versehen sind. Um diese Regeln zu befolgen, müssen Sie das illusorische Muster erkennen und sich an das erinnern, was Sie darüber wissen. Wenn Sie dies tun können, werden Sie sich nie mehr von der Müller-Lyer-Illusion hinters Licht führen lassen. Aber die eine Linie wird Ihnen nach wie vor länger erscheinen als die andere.

Es gibt nicht nur optische Täuschungen. Es gibt auch Illusionen des Denkens, die wir »kognitive Täuschungen« nennen. Als Student besuchte ich auch einige Lehrveranstaltungen über die Kunst und die Wissenschaft der Psychotherapie. In einer dieser Vorlesungen teilte uns der Dozent eine Lektion aus seiner langjährigen klinischen Erfahrung mit: »Hin und wieder werden Sie einem Patienten begegnen, der Sie mit der Äußerung verstören wird, dass in seinen zurückliegenden Therapien zahlreiche Fehler gemacht wurden. Er suchte mehrere klinische Therapeuten auf, und alle enttäuschten ihn. Der Patient kann in nachvollziehbarer Weise schildern, wie ihn seine Therapeuten missverstanden haben, aber er hat schnell bemerkt, dass Sie anders sind. Sie liegen auf der gleichen Wellenlänge, er ist überzeugt, dass Sie ihn verstehen und ihm helfen können.« An dieser Stelle hob der Dozent seine Stimme, als er sagte: »Kommen Sie bloß nicht auf den Gedanken, diesen Patienten zu nehmen! Werfen Sie ihn hinaus! Er ist wahrscheinlich ein Psychopath, und Sie werden ihm nicht helfen können.«

Viele Jahre später wurde mir klar, dass der Dozent uns vor dem Charme von Psychopathen gewarnt hatte, und der führende Experte auf dem Gebiet der Psychopathie-Forschung bestätigte mir, dass der Rat des Dozenten durchaus sinnvoll war.[5] Hier besteht eine große Ähnlichkeit mit der Müller-Lyer-Illusion. Uns wurde nicht beigebracht, welche Gefühle wir diesem Patienten entgegenbringen sollten. Unser Dozent ging einfach davon aus, dass das Mitgefühl, das wir für den Patienten empfinden würden, nicht von uns kontrolliert werden könne; es würde aus System 1 hervorgehen. Uns wurde auch nicht beigebracht, unseren Gefühlen für Patienten grundsätzlich zu misstrauen. Uns wurde gesagt, dass eine starke Sympathie für einen Patienten mit mehreren gescheiterten Therapien ein Gefahrenzeichen ist – wie die spitzen Winkel an den parallelen Linien. Es ist eine Illusion – eine kognitive Täuschung –, und mir (System 2) wurde beigebracht, wie man sie erkennt, und geraten, ihr nicht zu glauben oder nachzugeben.

Die Frage, die am häufigsten in Bezug auf kognitive Täuschungen gestellt wird, lautet: Kann man sie überwinden? Die Botschaft dieser Beispiele ist nicht ermutigend. Da System 1 automatisch operiert und nicht willentlich abgestellt werden kann, lassen sich intuitive Denkfehler oftmals nur schwer verhindern. Kognitive Verzerrungen lassen sich nicht immer vermeiden, weil System 2 vielleicht nichts von dem Fehler ahnt. Selbst wenn Hinweise auf wahrscheinliche Fehler vorliegen, lassen diese sich nur durch gesteigerte Überwachung und mühsame Aktivierung von System 2 verhüten. Im Alltagsleben aber ist beständige erhöhte Wachsamkeit nicht unbedingt gut, und sie ist zweifellos unpraktisch. Es wäre unerträglich mühsam, ständig sein eigenes Denken zu hinterfragen, und System 2 ist viel zu langsam und ineffizient, um bei Routine-Entscheidungen als ein Ersatz für System 1 zu fungieren. Wir können bestenfalls einen Kompromiss erreichen: lernen, Situationen zu erkennen, in denen Fehler wahrscheinlich sind, und uns stärker darum bemühen, weitreichende Fehler zu vermeiden, wenn viel auf dem Spiel steht. Die Prämisse dieses Buches lautet, dass man die Fehler anderer leichter erkennt als seine eigenen.

## Nützliche Fiktionen

Ich habe Sie gebeten, sich die beiden Systeme als mentale Akteure mit individuellen Persönlichkeiten, Fähigkeiten und Begrenzungen vorzustellen. Ich werde häufig Sätze verwenden, in denen die Systeme die Subjekte sind, wie etwa in »System 2 berechnet Produkte«.

Ein solcher Sprachgebrauch gilt in den Fachkreisen, in denen ich mich bewege, als eine Sünde, weil er die Gedanken und Handlungen einer Person durch die Gedanken und Handlungen kleiner Menschen im Kopf dieser Person zu erklären scheint.[6] Grammatikalisch gesehen, ähnelt der Satz über System 2 dem Satz: »Der Butler stiehlt die Portokasse.« Meine Kollegen würden darauf hinweisen, dass die Tat des Butlers das Verschwinden des Geldes erklärt, und sie fragen zu Recht, ob der Satz über System 2 erklärt, wie Produkte berechnet werden. Meine Antwort lautet, dass der kurze aktive Satz, der System 2 Berechnungen zuschreibt, eine Beschreibung, keine Erklärung sein soll. Er ist nur deshalb bedeutungsvoll, weil wir bereits etwas über System 2 wissen. Er ist eine Abkürzung des folgenden Satzes: »Mentale Arithmetik ist eine anstrengende willentliche Aktivität, die man nicht ausführen sollte, während man nach links abbiegt, und sie geht mit erweiterten Pupillen und beschleunigtem Herzschlag einher.«

In ähnlicher Weise bedeutet die Aussage, dass »das Fahren auf einer Fernstraße unter gewöhnlichen Umständen System 1 überlassen bleibt«, dass es ein automatisierter und fast müheloser Prozess ist, ein Auto durch eine Kurve zu steuern. Es bedeutet auch, dass ein routinierter Fahrer auf einer leeren Fernstraße fahren und gleichzeitig ein Gespräch führen kann. »System 2 verhinderte, dass James auf die Beleidigung in einer unklugen Weise reagierte« bedeutet schließlich, dass James aggressiver reagiert hätte, wenn seine Fähigkeit zu anstrengender willentlicher Kontrolle beeinträchtigt gewesen wäre (zum Beispiel, wenn er betrunken gewesen wäre).

System 1 und System 2 spielen in der Geschichte, die ich in diesem Buch erzähle, eine so zentrale Rolle, dass ich unmissverständlich klarmachen will, dass sie fiktive Figuren sind. System 1 und System 2 sind keine Systeme im üblichen Sinne, keine Gebilde aus Elementen oder Teilen, die miteinander wechselwirken. Und es gibt nicht einen bestimmten Teil des

Gehirns, in dem eines der beiden Systeme fest ansässig wäre. Sie mögen jetzt fragen: Wozu soll es gut sein, fiktive Figuren mit hässlichen Namen in einem Sachbuch einzuführen? Die Antwort lautet, dass diese Figuren aufgrund gewisser Eigenheiten des menschlichen Denkens nützlich sind. Ein Satz ist leichter zu verstehen, wenn er beschreibt, was ein Akteur (System 2) tut, als wenn er beschreibt, was etwas ist, welche Eigenschaften es aufweist. Anders gesagt, »System 2« ist ein besseres Subjekt für einen Satz als »mentale Arithmetik«. Unser Denkvermögen – insbesondere System 1 – scheint Geschichten über tatkräftige Akteure, die Persönlichkeiten, Gewohnheiten und Fähigkeiten besitzen, besonders gut konstruieren und interpretieren zu können. Wir haben uns sehr schnell eine schlechte Meinung über den diebischen Butler gebildet, wir erwarten, dass er weitere Missetaten begeht, und wir werden uns eine Zeit lang an ihn erinnern. Die gleiche Hoffnung hege ich in Bezug auf die Sprache der Systeme.

Weshalb nenne ich sie System 1 und System 2 anstatt »automatisches System« und »willentliches System«, was anschaulicher wäre? Das hat einen einfachen Grund: Es dauert länger, »automatisches System« auszusprechen als »System 1«, und aus diesem Grund nimmt dieser Ausdruck mehr Platz in Ihrem Arbeitsgedächtnis ein.[7] Das ist von Bedeutung, weil alles, was Ihr Arbeitsgedächtnis belegt, Ihr Denkvermögen vermindert. Sie sollten »System 1« und »System 2« als Spitznamen betrachten, ähnlich wie Bob und Joe, die Figuren darstellen, die Sie im Lauf der Lektüre dieses Buches kennenlernen werden. Die frei erfundenen Systembezeichnungen erleichtern es mir, über Urteils- und Entscheidungsprozesse nachzudenken, und sie erleichtern Ihnen das Verständnis meiner Darlegungen.

## Zum Thema »System 1 und System 2«

> »Er hatte einen Eindruck, aber einige seiner Eindrücke sind Illusionen.«

»Dies war eine reine System-1-Antwort. Sie reagierte auf die Bedrohung, bevor sie diese erkannte.«

»Hier spricht gerade Ihr System 1. Schalten Sie einen Gang zurück, und lassen Sie Ihr System 2 die Kontrolle übernehmen.«

# *2. Aufmerksamkeit und Anstrengung* [1]

In dem unwahrscheinlichen Fall, dass dieses Buch verfilmt werden sollte, wäre System 2 eine Nebenfigur, die sich für den Protagonisten hält. Die bestimmende Eigenschaft von System 2 in dieser Geschichte besteht darin, dass seine Operationen mit Anstrengung verbunden sind, und eines seiner Hauptmerkmale ist Faulheit, also die Neigung, nur die Mühe aufzuwenden, die absolut notwendig ist. Daher werden die Gedanken und Handlungen, von denen System 2 annimmt, dass es sie frei gewählt hat, oftmals von der Figur bestimmt, die im Mittelpunkt der Geschichte steht, System 1. Aber es gibt höchst wichtige Aufgaben, die nur System 2 erledigen kann, weil sie Anstrengung und Selbstbeherrschung erfordern, mit denen die Intuitionen und Impulse von System 1 überwunden werden.

## Mentale Anstrengung

Wenn Sie erleben möchten, wie es ist, wenn Ihr System 2 mit voller Kraft arbeitet, sollten Sie die folgende Übung machen; sie wird Sie innerhalb von fünf Sekunden an die Grenzen Ihrer kognitiven Leistungsfähigkeit bringen. Bilden Sie zunächst mehrere verschiedene vierstellige Ziffernfolgen, und schreiben Sie jede Folge auf eine Karteikarte. Legen Sie eine unbeschriebene Karte oben auf den Packen. Die Aufgabe, die Sie ausführen werden, heißt »Eins addieren«. Sie geht folgendermaßen:

> Beginnen Sie, einen regelmäßigen Rhythmus zu schlagen (oder, besser noch, stellen Sie ein Metrononom auf einen Schlag pro Sekunde ein). Nehmen Sie die leere Karte weg, und lesen Sie die vier Ziffern laut vor. Warten Sie zwei Schläge, und lesen Sie dann eine Ziffernfolge vor, in der die ursprünglichen Ziffern um eins hochgezählt werden. Wenn auf der Karte die Ziffern 5294 stehen, lautet die richtige Antwort 6305. Es ist wichtig, den Rhythmus zu halten.

Nur wenige Personen kommen bei der »Eins addieren«-Aufgabe mit mehr als vier Ziffern zurecht, aber wenn Sie eine schwierigere Herausforderung wollen, dann versuchen Sie es doch mal mit »Drei addieren«.

Wenn Sie wissen möchten, was Ihr Körper tut, während Ihr Gehirn schwer arbeitet, sollten Sie zwei Bücherstapel auf einem robusten Tisch aufbauen; auf einem davon platzieren Sie eine Videokamera, und auf den anderen stützen Sie Ihr Kinn auf. Schalten Sie die Videokamera ein, und starren Sie in die Kameralinse, während Sie die »Eins addieren«- oder »Drei addieren«-Übungen machen. Später finden Sie in der sich verändernden Größe Ihrer Pupillen ein getreuliches Protokoll ihrer mentalen Anstrengungen.

Ich habe eine lange persönliche Geschichte mit der »Eins addieren«-Übung. Zu Beginn meiner wissenschaftlichen Laufbahn verbrachte ich ein Jahr an der Universität Michigan als Gastwissenschaftler in einem Labor, in

dem Hypnose erforscht wurde. Auf der Suche nach einem sinnvollen Forschungsthema stieß ich auf einen Artikel im *Scientific American*, in dem der Psychologe Eckhard Hess die Pupille des Auges als Fenster zur Seele bezeichnete.[2] Vor Kurzem habe ich diesen Artikel noch einmal gelesen, und ich fand ihn abermals inspirierend. Er beginnt damit, dass Hess berichtet, seine Frau habe bemerkt, wie sich seine Pupillen weiteten, als er schöne Naturaufnahmen betrachtete, und er endet mit zwei bemerkenswerten Bildern von derselben gut aussehenden Frau, die auf dem einen Bild viel attraktiver erscheint als auf dem anderen. Dabei gibt es nur einen Unterschied: Auf dem attraktiven Bild scheinen die Pupillen geweitet, auf dem anderen verengt zu sein. Hess schrieb auch über Belladonna, eine pupillenvergrößernde Substanz, die früher als Kosmetikartikel verwendet wurde, und von Basarkäufern, die dunkle Brillen tragen, um die Stärke ihres Interesses vor den Händlern zu verbergen. Vor allem einer der Befunde von Hess erregte meine Aufmerksamkeit. Ihm war aufgefallen, dass die Pupillen empfindliche Indikatoren der mentalen Anstrengung sind – sie vergrößern sich stärker, wenn die Probleme schwierig sind, als wenn sie leicht sind. Seine Beobachtungen deuteten darauf hin, dass sich die Reaktion auf mentale Anstrengung von der auf emotionale Erregung unterscheidet. Seine Arbeit hatte nicht viel mit Hypnose zu tun, aber ich gelangte zu dem Schluss, dass die Hypothese, dass es einen sichtbaren Indikator für geistige Anstrengung gibt, ein vielversprechendes Forschungsthema ist. Jackson Beatty, ein Doktorand im Labor, teilte meine Begeisterung, und so machten wir uns an die Arbeit.

Beatty und ich entwarfen eine Versuchsanordnung, bei der wir einen Raum so ähnlich gestalteten wie den Untersuchungsraum eines Optikers: Die Versuchspersonen lehnten ihren Kopf auf eine Kinn- und Kopfstütze und starrten in eine Kamera, während sie aufgezeichneten Informationen lauschten und Fragen im Rhythmus der aufgezeichneten Schläge eines Metronoms beantworteten. Die Schläge lösten jede Sekunde einen Infrarotblitz aus, wobei gleichzeitig eine Aufnahme gemacht wurde. Am Ende jedes Versuchsdurchgangs entwickelten wir so schnell wie möglich den Film, projizierten die Bilder der Pupille auf eine Leinwand und vermaßen sie mit einem Lineal. Die Methode eignete sich hervorragend für junge, ungeduldige Forscher: Wir kannten unsere Ergebnisse praktisch sofort, und sie waren immer eindeutig.

Beatty und ich konzentrierten uns auf schnelle Aufgaben, wie etwa »Eins addieren«, bei denen wir genau wussten, was die Versuchsperson zu jedem beliebigen Zeitpunkt dachte.[3] Wir zeichneten Ziffernfolgen zu den Schlägen des Metronoms auf und wiesen die Versuchsperson an, die Ziffern nacheinander zu wiederholen oder zu verändern und dabei den gleichen Rhythmus beizubehalten. Wir fanden schon bald heraus, dass die Pupillengröße sekündlich schwankte, worin sich die unterschiedlichen Anforderungen der Aufgabe widerspiegelten. Die Antwortkurve hatte die Form eines auf dem Kopf stehenden V. Wie Sie es selbst erlebt haben, wenn Sie »Eins addieren« und »Drei addieren« ausprobierten, wächst die Anstrengung mit jeder zusätzlichen Ziffer, die man hört, erreicht einen unerträglichen Höhepunkt, wenn man sich beeilt, während und unmittelbar nach der Pause eine umgewandelte Ziffernfolge zu erzeugen, und lässt allmählich nach, wenn man sein Kurzzeitgedächtnis »entlädt«. Die Pupillendaten entsprachen exakt dem subjektiven Erleben: Längere Ziffernfolgen verursachten durchweg eine stärkere Vergrößerung, die Umwandlungsaufgabe erhöhte die Anstrengung, und die Pupillengröße erreichte ihr Maximum bei der stärksten mentalen Anstrengung. »Eins addieren« mit vier Ziffern führte zu einer stärkeren Pupillenvergrößerung als die Aufgabe, sich sieben Ziffern zum sofortigen Abruf zu merken. »Drei addieren«, das sehr viel schwerer ist, ist die anspruchsvollste Aufgabe, die ich je gesehen habe. In den ersten fünf Sekunden vergrößert sich die Pupille um etwa 50 Prozent ihrer ursprünglichen Ausdehnung, und der Herzschlag erhöht sich um etwa sieben Schläge pro Minute.[4] Hier stoßen Menschen an ihre Belastungsgrenze – sie geben auf, wenn von ihnen noch mehr verlangt wird. Als wir unseren Probanden mehr Ziffern vorführten, als sie sich merken konnten, vergrößerten sich ihre Pupillen nicht mehr, oder sie schrumpften sogar.

Wir arbeiteten einige Monate lang in großzügigen Räumlichkeiten, in denen wir ein Videosystem installiert hatten, das ein Bild von der Pupille der Versuchsperson auf eine Leinwand im Flur projizierte; wir konnten auch hören, was im Labor geschah. Der Durchmesser der projizierten Pupille betrug etwa einen Fuß; zu sehen, wie sie sich erweiterte und zusammenzog, wenn der Proband bei der Arbeit war, war ein faszinierender Anblick, eine ziemliche Attraktion für Besucher in unserem Labor. Wir amüsierten uns und beeindruckten unsere Gäste durch unsere Fähigkeit, zu erahnen, wann

der Proband aufgeben würde. Bei einer mentalen Multiplikation erweiterte sich die Pupille normalerweise innerhalb weniger Sekunden sehr stark und blieb vergrößert, solange die Person an dem Problem arbeitete; sobald sie eine Lösung fand oder aufgab, zog sich die Pupille sofort zusammen. Wenn wir vom Flur aus zusahen, überraschten wir manchmal den Besitzer der Pupille und unsere Gäste mit der Frage: »Weshalb haben Sie gerade jetzt aufgegeben?« Die Antwort aus dem Labor lautete oft: »Woher wissen Sie das?«, worauf wir entgegneten: »Wir haben ein Fenster zu Ihrer Seele.«

Die beiläufigen Beobachtungen, die wir vom Flur aus machten, waren manchmal so informativ wie die förmlichen Experimente. Ich machte eine wichtige Entdeckung, als ich, während einer Pause zwischen zwei Aufgaben, die Pupille einer Frau betrachtete. Sie lehnte den Kopf weiterhin gegen die Kinnstütze, sodass ich das Bild ihres Auges sehen konnte, während sie ein Gespräch mit dem Experimentator anknüpfte. Ich war erstaunt, zu sehen, dass die Pupille klein blieb und sich nicht merklich vergrößerte, als sie sprach und zuhörte. Anders als die Aufgaben, die wir erforschten, erforderte das triviale Gespräch offenbar nur wenig oder gar keine Anstrengung – nicht mehr, als man für das Abspeichern von zwei oder drei Ziffern benötigt. Das war ein Aha-Erlebnis: Mir wurde klar, dass die von uns ausgewählten Aufgaben ungewöhnlich anstrengend waren. Mir kam ein Bild in den Sinn: Denkprozesse – heute würde ich von System-2-Prozessen sprechen – laufen normalerweise im Tempo eines gemütlichen Spaziergangs ab, der hin und wieder durch Jogging-Etappen und selten durch einen rasanten Sprint unterbrochen wird. Die »Eins addieren«- und »Drei addieren«-Übungen sind Sprints, und das beiläufige Plaudern ist ein Spaziergang.

Wir fanden heraus, dass Menschen, die einen mentalen Sprint laufen, faktisch blind werden können. Die Autoren des Buches *Der unsichtbare Gorilla* hatten den Gorilla »unsichtbar« gemacht, indem sie die Aufmerksamkeit der Betrachter mit der Aufgabe, die Ballwechsel zu zählen, intensiv beanspruchten. Wir berichteten von einem weit weniger spektakulären Beispiel von Blindheit während einer »Eins addieren«-Übung. Wir boten unseren Probanden eine Reihe rasch aufblinkender Buchstaben dar, während sie arbeiteten.[5] Wir sagten ihnen, die Lösung der Aufgabe habe absoluten Vorrang, aber sie wurden auch gebeten, am Ende der Ziffernaufgabe anzugeben, ob irgendwann während des Versuchs der Buchstabe K aufgetaucht sei. Wir fanden vor allem heraus, dass sich die

Fähigkeit, den gesuchten Buchstaben wahrzunehmen und zu melden, im Verlauf der zehn Sekunden dauernden Übung veränderte. Den Beobachtern entging fast nie ein K, das am Anfang oder gegen Ende der »Eins addieren«-Aufgabe gezeigt wurde, aber der gesuchte Buchstabe entging ihnen in fast 50 Prozent der Fälle, wenn die mentale Anstrengung ihren Höhepunkt erreichte, obwohl die Bilder verrieten, dass sie mit weit aufgerissenen Augen direkt auf den Buchstaben schauten. Die Nichtwahrnehmung folgte dem gleichen umgekehrten V-Muster wie die Pupillenvergrößerung. Die Ähnlichkeit war beruhigend: Die Pupillenweite war ein gutes Maß des *arousal*, des körperlichen Aktivierungszustands, der mit mentaler Anstrengung verbunden ist, und so konnten wir weitermachen und sie zur Erforschung kognitiver Prozesse nutzen.

Ganz ähnlich wie der Stromzähler außerhalb Ihres Hauses oder Ihrer Wohnung geben die Pupillen zuverlässig Aufschluss über die laufende Verbrauchsrate mentaler Energie.[6] Diese Analogie geht sehr weit. Ihr Stromverbrauch hängt davon ab, was Sie tun – ob Sie ein Zimmer beleuchten oder eine Scheibe Brot toasten. Wenn Sie eine Glühbirne oder einen Toaster anschalten, verbrauchen diese die Energie, die sie benötigen, aber kein bisschen mehr. In ähnlicher Weise entscheiden auch wir, was wir tun, aber wir haben nur beschränkte Kontrolle über die Anstrengung, die dazu erforderlich ist. Angenommen, Ihnen werden vier Ziffern gezeigt, zum Beispiel 9462, und man sagt Ihnen, dass Ihr Leben davon abhängt, dass Sie sich diese Ziffern zehn Sekunden lang merken. Sosehr Sie auch am Leben hängen, können Sie für diese Aufgabe nicht so viel Mühe aufwenden, wie Sie aufbringen müssten, um zu denselben Ziffern jeweils drei hinzuzählen.

System 2 und die Stromkreise in Ihrem Haus besitzen eine begrenzte Kapazität, aber sie reagieren unterschiedlich auf eine drohende Überlastung. Ein Stromkreisunterbrecher wird aktiviert, wenn der Strombedarf eine kritische Schwelle überschreitet, was dazu führt, dass die Stromversorgung aller Geräte an diesem Stromkreis gleichzeitig unterbrochen wird. Dagegen ist die Reaktion auf mentale Überlastung selektiv und präzise: System 2 schützt die wichtigste Aktivität, sodass sie die Aufmerksamkeit erhält, die sie braucht; »freie Kapazität« wird im Sekundenrhythmus anderen Aufgaben zugewiesen. In unserer Version des Gorilla-Experiments wiesen wir die Teilnehmer an, der Ziffernaufgabe Priorität einzuräumen. Wir wissen, dass sie dieser Anweisung Folge leisteten, weil der Zeitpunkt, zu dem der optische Stimulus präsentiert

wurde, keinen Einfluss auf die Hauptaufgabe hatte. Wenn der kritische Buchstabe zu einer Zeit hoher Aufmerksamkeitsbeanspruchung dargeboten wurde, sahen ihn die Probanden schlichtweg nicht. Wenn die rechnerische Umformungsaufgabe weniger anspruchsvoll war, war die Erkennungsrate höher.

Das System der hochdifferenzierten Aufmerksamkeitszuweisung wurde über lange evolutionäre Zeiträume immer weiter ausgefeilt. Es erhöhte die Überlebenschancen, wenn man die schwerwiegendsten Bedrohungen oder die vielversprechendsten Gelegenheiten schnell erkannte und umgehend darauf reagierte, und diese Fähigkeit ist zweifellos nicht auf Menschen beschränkt. Selbst beim modernen Menschen übernimmt System 1 in Notlagen die Kontrolle und weist Handlungen, die dem Selbstschutz dienen, höchste Priorität zu. Stellen Sie sich vor, Sie sitzen am Steuer eines Wagens, der unerwartet auf einem großen Ölfleck ins Rutschen gerät. Sie werden feststellen, dass Sie auf die Gefahr reagierten, bevor sie Ihnen voll und ganz bewusst wurde.

Beatty und ich arbeiteten nur ein Jahr zusammen, aber unsere Zusammenarbeit hatte weitreichende Folgen für unsere weiteren wissenschaftlichen Laufbahnen. Er wurde schließlich zum führenden Experten auf dem Gebiet der »kognitiven Pupillometrie«, und ich schrieb ein Buch mit dem Titel *Attention and Effort*, das zu einem großen Teil auf unseren gemeinsamen Erkenntnissen und anschließenden Forschungsarbeiten, die ich im Jahr darauf in Harvard durchführte, basierte. Wir lernten eine Menge über kognitive Funktionsmechanismen – die ich heute unter dem Oberbegriff »System 2« zusammenfassen würde –, indem wir bei unterschiedlichsten Aufgaben die Pupillengrößen unserer Probanden maßen.

Je mehr Geschick man bei der Lösung einer Aufgabe entwickelt, umso weniger Energie muss man für sie aufwenden. Gehirnscans von Versuchspersonen haben gezeigt, dass sich das mit einer Handlung verbundene Aktivitätsmuster mit der Fertigkeit verändert, da weniger Gehirnregionen daran beteiligt sind.[7] Begabung hat ähnliche Wirkungen. Hochintelligente Menschen lösen die gleichen Probleme müheloser, wie sowohl die Pupillengröße als auch die Hirnaktivität anzeigen.[8] Ein allgemeingültiges »Gesetz des geringsten Aufwands« gilt sowohl für kognitive als auch für physische Anstrengungen.[9] Das Gesetz besagt, dass

Menschen dann, wenn es mehrere Wege gibt, um das gleiche Ziel zu erreichen, schließlich den Weg wählen, der mit dem geringsten Arbeitsaufwand verbunden ist. In der Ökonomie der Handlungen ist die Anstrengung ein Kostenfaktor, und hinter dem Erwerb von Fähigkeiten steht das Streben nach einem ausgewogenen Kosten-Nutzen-Verhältnis.[10] Faulheit ist tief in unserer Natur angelegt.

Die von uns untersuchten Aufgaben hatten sehr unterschiedliche Auswirkungen auf die Pupillengröße. Zu Beginn des Versuchs waren unsere Probanden wach, aufmerksam und bereit, sich einer Aufgabe zuzuwenden – vermutlich auf einem höheren Niveau des *arousal* und der kognitiven Bereitschaft als gewöhnlich. Wenn sich die Probanden ein oder zwei Ziffern merken oder lernen sollten, ein Wort mit einer Ziffer zu assoziieren (3 = Tür), führte dies durchgängig zu einer kurzzeitigen Erhöhung des a*rousal* über das Ausgangsniveau, aber die Effekte waren sehr gering – bei einer »Drei addieren«-Aufgabe nahm der Pupillen-Durchmesser nur um 5 Prozent zu. Bei einer Aufgabe, bei der es darum ging, zwei Tonhöhen zu unterscheiden, fiel die Pupillenerweiterung deutlich größer aus. Jüngste Forschungen haben gezeigt, dass die Hemmung der Tendenz, ablenkende Wörter zu lesen (wie in Abbildung 2 des vorangehenden Kapitels), ebenfalls eine gewisse Anstrengung erfordert.[11] Tests des Kurzzeitgedächtnisses mit sechs oder sieben Ziffern waren anstrengender. Wie Sie selbst erleben können, erfordert auch die Bitte, sich an Ihre Telefonnummer oder den Geburtstag Ihres Ehepartners zu erinnern und diese laut auszusprechen, eine kurze, aber erhebliche Anstrengung, weil die ganze Ziffernfolge im Gedächtnis gehalten werden muss, während eine Antwort organisiert wird. Die mentale Multiplikation zweistelliger Zahlen und die »Drei addieren«-Aufgabe liegen an der kognitiven Leistungsgrenze der meisten Menschen.

Was macht manche kognitiven Aktivitäten anspruchsvoller und anstrengender als andere? Welche Ergebnisse müssen wir mit der Währung der Aufmerksamkeit kaufen? Was kann System 2, was System 1 nicht kann? Wir haben jetzt vorläufige Antworten auf diese Fragen.

Anstrengung ist erforderlich, um mehrere Vorstellungen, die verschiedene Handlungen erfordern oder die gemäß einer Regel kombiniert werden müssen, gleichzeitig im Gedächtnis zu halten – beim Betreten des Supermarktes die Einkaufsliste in Gedanken durchgehen, in einem

Restaurant zwischen Fisch und Kalbfleisch wählen oder ein überraschendes Ergebnis einer Erhebung mit der Information verbinden, dass die Stichprobe zu klein war, zum Beispiel. Nur System 2 kann Regeln befolgen, Objekte in Bezug auf mehrere Merkmale vergleichen und wohlüberlegte Wahlen zwischen Optionen treffen. Das automatische System 1 besitzt diese Fähigkeiten nicht. System 1 erkennt einfache Beziehungen (»Sie sehen sich alle ähnlich«, »Der Sohn ist viel größer als der Vater«) und kann Informationen über einen Gegenstand hervorragend zusammenführen, aber es kann nicht mehrere verschiedene Aufgaben gleichzeitig bewältigen, und auch mit rein statistischen Informationen kann es wenig anfangen. System 1 erkennt, dass eine Person, die als »sanftmütiger und ordentlicher Mensch, der Ordnung und Struktur braucht und Liebe zum Detail besitzt«, beschrieben wird, der Karikatur eines Bibliothekars ähnelt, aber nur System 2 kann die Aufgabe ausführen, diese Intuition mit dem Wissen über die geringe Anzahl von Bibliothekaren zu verknüpfen – sofern System 2 weiß, wie es dies tun muss, was nur für wenige Menschen gilt.

Eine zentrale Fähigkeit von System 2 ist die Einführung von *task sets* (kognitive Zustände der optimalen Aufgabenvorbereitung): Es kann das Gedächtnis so programmieren, dass es einer Anweisung gehorcht, die die üblichen Reaktionen außer Kraft setzt. Nehmen wir folgendes Beispiel: Zählen Sie, wie oft der Buchstabe f auf dieser Seite vorkommt. Diese Aufgabe haben Sie noch nie verrichtet, und sie wird Ihnen nicht leichtfallen, aber Ihr System 2 kann sie übernehmen. Es wird Sie Mühe kosten, sich auf diese Übung vorzubereiten, und es wird mühsam sein, sie auszuführen, auch wenn Sie mit zunehmender Geübtheit zweifellos besser werden. Psychologen sprechen bezüglich der Aktivierung und Beendigung von *task sets* von »exekutiver Kontrolle«, und Neurowissenschaftler haben die Hauptregionen des Gehirns identifiziert, von denen diese exekutiven Funktionen getragen werden. Eine dieser Regionen ist immer beteiligt, wenn ein Konflikt gelöst werden muss. Eine andere ist der präfrontale Cortex, eine Region, die beim Menschen viel stärker entwickelt ist als bei anderen Primaten, und sie ist an Aktivitäten beteiligt, die wir mit Intelligenz assoziieren. [12]

Nehmen wir jetzt an, dass Sie am Ende der Seite eine weitere Anweisung erhalten: Zählen Sie alle Kommata auf der nächsten Seite. Dies wird Ihnen schwererfallen, weil sie die neu erworbene Tendenz überwinden müssen, die Aufmerksamkeit auf den Buchstaben f zu fokussieren. Zu den

wichtigsten Entdeckungen, die in den letzten Jahrzehnten in der Kognitionspsychologie gemacht wurden, gehört die Tatsache, dass der Aufgabenwechsel mühsam ist, insbesondere unter Zeitdruck.[13] Das Erfordernis einer raschen Umstellung ist einer der Gründe dafür, dass »Drei addieren« und mentale Multiplikation so schwierig sind. Bei der »Drei addieren«-Aufgabe müssen Sie mehrere Ziffern gleichzeitig in Ihrem Arbeitsgedächtnis halten und jede mit einer bestimmten Operation assoziieren: Einige Ziffern warten darauf, umgeformt zu werden, eine wird gerade umgewandelt, und andere, die bereits umgeformt wurden, werden zum Zweck der Berichterstattung vorgehalten.[14] Bei modernen Tests des Arbeitsgedächtnisses sollen die Probanden wiederholt zwischen zwei anspruchsvollen Aufgaben wechseln und sich die Ergebnisse der einen Operation merken, während sie die andere ausführen. Personen, die bei diesen Tests gut abschneiden, schneiden tendenziell auch bei Tests der allgemeinen Intelligenz gut ab.[15] Doch die Fähigkeit zur Aufmerksamkeitssteuerung ist nicht bloß ein Maß der Intelligenz; Maße für die Effizienz der Aufmerksamkeitssteuerung sagen die Leistungsfähigkeit von Fluglotsen und Piloten der israelischen Luftwaffe mit größerer Zuverlässigkeit vorher als der IQ.[16]

Auch Zeitdruck steigert die notwendigen Anstrengungen. Als Sie die »Drei addieren«-Übung ausführten, ergab sich der Zeitdruck teils durch das Metronom und teils durch die Gedächtnisbelastung. Wie jemand, der mit mehreren Bällen in der Luft jongliert, können Sie es sich nicht leisten, langsamer zu werden; die Schnelligkeit, mit der sich Gedächtnisinhalte verflüchtigen, zwingt zur Eile: Sie müssen Informationen auffrischen und wiederholen, ehe sie verloren gehen. Jede Aufgabe, die von Ihnen verlangt, mehrere Gedanken gleichzeitig im Sinn zu behalten, zwingt Sie zu schnellem Handeln. Wenn Sie nicht das Glück haben, über ein großes Arbeitsgedächtnis zu verfügen, sind Sie vielleicht gezwungen, unangenehm hart zu arbeiten. Die anstrengendsten Formen langsamen Denkens sind diejenigen, die von Ihnen verlangen, schnell zu denken.

Bei der »Drei addieren«-Aufgabe ist Ihnen bestimmt aufgefallen, wie ungewohnt es für Sie ist, sich mental derartig anzustrengen. Selbst wenn Sie Ihren Lebensunterhalt mit geistiger Arbeit bestreiten, sind nur wenige der mentalen Aufgaben, die Sie im Lauf eines Arbeitstags erledigen, so anspruchsvoll wie »Drei addieren« oder auch wie das Abspeichern von

sechs Ziffern zum sofortigen Abruf. Wir vermeiden mentale Überlastung normalerweise dadurch, dass wir unsere Aufgaben in mehrere leichte Schritte unterteilen, Zwischenergebnisse im Langzeitgedächtnis ablegen oder notieren, statt sie einem allzu leicht überlasteten Arbeitsgedächtnis anzuvertrauen. Wir überbrücken große Entfernungen, indem wir uns Zeit lassen, und wir halten uns beim Denken an das Gesetz der geringsten Anstrengung.

## Zum Thema »Aufmerksamkeit und Anstrengung«

»Ich werden nicht versuchen, dies beim Fahren zu lösen. Dies ist eine pupillenvergrößernde Aufgabe. Sie erfordert mentale Anstrengung!«

»Hier gilt das Gesetz vom geringsten Aufwand. Er wird so wenig wie möglich nachdenken.«

»Sie hat die Sitzung nicht vergessen. Sie war voll und ganz auf etwas anderes konzentriert, als die Sitzung vereinbart wurde, und sie hat Sie einfach nicht gehört.«

»Mir fiel sofort eine Intuition von System 1 ein. Ich muss von Neuem beginnen und mein Gedächtnis gezielt durchsuchen.«

# *3. Der faule Kontrolleur*

Ich verbringe jedes Jahr einige Monate in Berkeley, und eine meiner größten Freuden dort ist ein täglicher Spaziergang über sechseinhalb Kilometer auf einem ausgewiesenen Weg in den Hügeln, mit einem großartigen Blick auf die Bucht von San Francisco. Ich achte normalerweise auf meine Zeit, und das hat mich eine Menge über Anstrengung gelehrt. Ich habe eine für mich angenehme Geschwindigkeit gefunden, etwa 17 Minuten für etwa eineinhalb Kilometer, was ich als gemütliches Gehtempo erlebe. Ich strenge mich körperlich an und verbrenne bei dieser Geschwindigkeit gewiss mehr Kalorien, als wenn ich in einem Lehnstuhl sitzen würde, aber ich erlebe keinen Druck, keinen Konflikt und auch nicht das Bedürfnis, mich selbst anzutreiben. Wenn ich in diesem Tempo gehe, kann ich nachdenken und arbeiten. Tatsächlich vermute ich, dass die leichte körperliche Aktivierung durch den Spaziergang zu einer größeren geistigen Agilität führt.

System 2 hat ebenfalls eine natürliche Geschwindigkeit. Man wendet ein wenig mentale Energie für zufällige Gedanken auf und dafür, zu erfassen, was um einen herum vor sich geht, auch wenn man geistig nicht auf eine bestimmte Aufgabe konzentriert ist, aber die Beanspruchung ist gering. Sofern man sich nicht in einer Situation befindet, die einen ungewöhnlich wachsam oder verlegen macht, muss man sich nicht besonders anstrengen, um zu erfassen, was in der Umgebung oder im Kopf geschieht. Man trifft viele kleine Entscheidungen, wenn man Auto fährt, beim Lesen der Zeitung Informationen aufnimmt und mit einem Partner oder Kollegen Artigkeiten austauscht – dies alles mit geringer Anstrengung und ohne Druck. Wie bei einem Spaziergang.

Es ist normalerweise leicht und sogar recht angenehm, spazieren zu gehen und gleichzeitig nachzudenken, aber im Extremfall scheinen diese Aktivitäten um die begrenzten Ressourcen von System 2 zu konkurrieren. Sie können diese Behauptung durch ein einfaches Experiment selbst überprüfen. Während Sie in gemütlichem Tempo mit einem Freund spazieren gehen, bitten Sie ihn, 23 × 78 im Kopf zu berechnen, und zwar sofort. Er wird höchstwahrscheinlich unvermittelt stehen bleiben. Bei mir

ist es so, dass ich beim Spazieren zwar nachdenken, aber keine mentale Arbeit verrichten kann, die das Kurzzeitgedächtnis stark beansprucht. Wenn ich eine komplizierte Beweisführung unter Zeitdruck entwickeln muss, würde ich mich lieber nicht bewegen, und ich würde lieber sitzen als stehen. Natürlich erfordert nicht jedes langsame Denken diese Form von intensiver Konzentration und anstrengender Berechnung – die besten Einfälle meines Lebens hatte ich auf gemütlichen Spaziergängen mit Amos. Wenn ich nicht mehr schlendere, sondern einen Schritt zulege, verändert dies völlig mein Erleben des Spazierens, weil der Wechsel in eine schnellere Gangart zu einer deutlichen Verschlechterung meiner Fähigkeit führt, zusammenhängend zu denken. Wenn ich beschleunige, richtet sich meine Aufmerksamkeit immer häufiger auf die Erfahrung des Gehens und auf die absichtliche Aufrechterhaltung eines höheren Tempos. Meine Fähigkeit, einen Gedankengang zu einem Abschluss zu bringen, ist entsprechend beeinträchtigt. Bei der höchsten Geschwindigkeit, die ich in dem hügeligen Gelände durchhalten kann, etwa 14 Minuten für 1,6 Kilometer, versuche ich erst gar nicht, an etwas anderes zu denken. Zusätzlich zu der physischen Anstrengung, die damit verbunden ist, meinen Körper zügig fortzubewegen, ist eine mentale Anstrengung der Selbstkontrolle erforderlich, um dem Impuls zu widerstehen, langsamer zu gehen. Selbstkontrolle und bewusstes Denken schöpfen anscheinend aus dem gleichen begrenzten Budget mentaler Arbeitskraft.

Die Aufrechterhaltung einer zusammenhängenden Gedankenführung und gelegentliches anstrengendes Nachdenken verlangen von den meisten von uns in aller Regel ebenfalls Selbstkontrolle. Obgleich ich keine systematische Studie durchgeführt habe, vermute ich, dass häufiger Aufgabenwechsel und beschleunigte mentale Arbeit nicht per se angenehm sind und dass Menschen sie, wenn möglich, vermeiden. So wird das Gesetz der geringsten Anstrengung zu einem Gesetz. Selbst bei fehlendem Zeitdruck erfordert die Aufrechterhaltung einer kohärenten Gedankenführung Disziplin. Jemand, der mir dabei zusähe, wie oft ich während einer Stunde, in der ich schreibe, in meinem E-Mail-Account nachsehe oder den Kühlschrank inspiziere, könnte zu dem nachvollziehbaren Schluss gelangen, dass ich mich gern ablenke und dass es mehr Selbstkontrolle von mir verlangt, als ich problemlos aufbringen kann, diesem Impuls zu widerstehen.

Glücklicherweise löst kognitive Arbeit nicht immer die Tendenz aus, ihr auszuweichen, und manchmal strengen sich Menschen über längere Zeiträume intensiv an, ohne dafür Willenskraft aufwenden zu müssen. Der Psychologe Mihaly Csikszentmihalyi (ausgesprochen wie »Six-Cent-Mihaly«) hat mehr als jeder andere getan, um diesen Zustand der mühelosen geistigen Versenkung zu erforschen, und der Name, den er dafür vorschlug, »Flow«, hat sogar Eingang in die Alltagssprache gefunden. Menschen, die einen Flow erleben, beschreiben diesen als »einen Zustand der mühelosen Konzentration, der so tief ist, dass sie ihr Gefühl für die Zeit, für sich selbst und für ihre Probleme verlieren«, und ihre Beschreibungen der Freude, die sie in diesem Zustand empfinden, sind so bezwingend, dass Csikszentmihalyi den Flow als eine »optimale Erfahrung« bezeichnet hat.[1] Viele Aktivitäten können ein Gefühl des Flows auslösen, angefangen vom Malen bis zur Teilnahme an Motorradrennen – und für einige glückliche Autoren, die ich kenne, ist selbst das Schreiben eines Buches oftmals eine optimale Erfahrung. Flow trennt fein säuberlich zwischen den beiden Formen der mentalen Anstrengung: Konzentration auf einen Aufgabe und bewusste Aufmerksamkeitssteuerung. Ein Motorrad mit 240 Stundenkilometern fahren und an einer Schachmeisterschaft teilnehmen sind zweifellos mental sehr anstrengende Tätigkeiten. In einem Flow-Zustand jedoch erfordert die Aufrechterhaltung der auf diese absorbierenden Aktivitäten fokussierten Aufmerksamkeit keinen Akt der Selbstkontrolle, wodurch Ressourcen freigesetzt werden, die für die anstehende Aufgabe verwendet werden können.

## Das ausgelastete und erschöpfte System 2

Es ist heute gängige Lehrmeinung, dass sowohl Selbstkontrolle als auch kognitive Anstrengung Formen mentaler Arbeit sind. Etliche psychologische Studien haben gezeigt, dass Menschen, die gleichzeitig mit einer anspruchsvollen kognitiven Aufgabe und mit einer Versuchung konfrontiert sind, eher der Versuchung nachgeben. Stellen Sie sich vor, Sie sollen sich eine Liste mit sieben Ziffern eine oder zwei Minuten lang einprägen. Man sagt Ihnen, das Einprägen der Ziffern habe für Sie oberste Priorität. Während Ihre Aufmerksamkeit auf die Ziffern gerichtet ist, werden Ihnen zwei Nachspeisen angeboten: ein sündhafter Schokoladenkuchen und ein tugendsamer Obstsalat. Die empirischen Befunde sprechen dafür, dass Sie eher den verlockenden Schokoladenkuchen auswählen werden, wenn Ihr Gehirn mit Ziffern beschäftigt ist. System 1 hat mehr Einfluss auf das Verhalten, wenn System 2 beschäftigt ist, und es hat eine Schwäche für Süßes.[2]

Menschen, die *kognitiv ausgelastet* sind, treffen auch eher egoistische Entscheidungen, verwenden sexistische Ausdrücke und fällen in sozialen Situationen oberflächliche Urteile.[3] Das Auswendiglernen und Aufsagen von Ziffern lockert die Kontrolle von System 2 über das Verhalten, aber natürlich ist die kognitive Belastung nicht der einzige Grund für eine geschwächte Selbstkontrolle. Ein paar Drinks haben die gleiche Wirkung, ebenso eine schlaflose Nacht. Die Selbstkontrolle von Morgenmenschen ist nachts beeinträchtigt; das Umgekehrte gilt für Nachtmenschen. Wenn man sich allzu viele Gedanken darüber macht, wie gut man eine Aufgabe erledigt, beeinträchtigt dies manchmal die Leistungsfähigkeit, weil das Kurzzeitgedächtnis von sinnlosen und sorgenvollen Gedanken überflutet wird.[4] Die Schlussfolgerung ist einfach: Selbstkontrolle erfordert Aufmerksamkeit und Anstrengung. Man kann es auch anders formulieren: Gedanken und Verhaltensweisen zu kontrollieren ist eine der Aufgaben, die System 2 ausführt.

Der Psychologe Roy Baumeister und seine Kollegen haben in einer Reihe überraschender Experimente den schlüssigen Nachweis erbracht, dass alle Spielarten willentlicher Anstrengung – kognitive, emotionale oder

physische – zumindest teilweise aus einem gemeinsamen Pool mentaler Energie schöpfen. Bei ihren Experimenten haben sie sukzessive und keine simultanen Aufgaben verwendet.

Baumeisters Gruppe hat wiederholt festgestellt, dass Willensanstrengung oder Selbstkontrolle ermüdend ist; wenn man sich zu einer Handlung zwingen muss, ist man weniger gewillt oder imstande, Selbstkontrolle auszuüben, wenn sich die nächste Herausforderung stellt. Das Phänomen wird »Ego-Depletion« (»Selbsterschöpfung«) genannt. In einem typischen Experiment schneiden Versuchspersonen, die aufgefordert werden, ihre emotionale Reaktion auf einen emotional aufgeladenen Film zu unterdrücken, später bei einem Test ihrer körperlichen Ausdauer – wie lange können sie trotz wachsenden Unbehagens einen Kraftmesser fest im Griff behalten – schlecht ab. Die emotionale Anstrengung in der ersten Phase des Experiments verringert die Fähigkeit, die Schmerzen anhaltender Muskelkontraktion zu ertragen, und aus diesem Grund erliegen »selbsterschöpfte« Menschen eher dem Impuls, aufzugeben. In einem anderen Experiment wird die selbstregulatorische Energie der Versuchspersonen zunächst durch eine Aufgabe erschöpft, bei der sie tugendhafte Nahrungsmittel wie Rettich und Sellerie verspeisen, während sie der Versuchung widerstehen, sich an Schokolade und fettreichen Keksen gütlich zu tun. Später werden diese Personen dann bei einer schwierigen kognitiven Aufgabe eher als üblich aufgeben. Die Liste der Situationen und Aufgaben, die bekanntermaßen die Selbstkontrolle erschöpfen, ist lang und vielfältig. Bei allen geht es um Konflikte und die Notwendigkeit, eine natürliche Neigung zu unterdrücken. Dazu gehören:

Nicht an weiße Bären denken
Die emotionale Reaktion auf einen aufwühlenden Film hemmen
Eine Reihe von konfliktbeladenen Entscheidungen treffen
Andere beeindrucken wollen
Freundlich auf das schlechte Verhalten eines Partners reagieren
Mit einem Menschen anderer Hautfarbe in Austausch treten (für voreingenommene Personen)

Die Liste der Hinweise auf erschöpfte Selbstkontrolle ist ebenfalls höchst unterschiedlich:

Von seiner Ernährung abweichen
Zu viel Geld für Impulskäufe ausgeben
Aggressiv auf Provokation reagieren
Bei einer Kraftmesser-Aufgabe weniger lange durchhalten
Bei Denkaufgaben und logischer Entscheidungsfindung schlecht abschneiden

Die Datenlage ist eindeutig: Aktivitäten, die hohe Anforderungen an System 2 stellen, erfordern Selbstkontrolle, und die Ausübung von Selbstkontrolle ist erschöpfend und unangenehm.[5] Im Unterschied zur kognitiven Belastung ist die Ego-Depletion zumindest teilweise ein Motivationsverlust. Nach der Ausübung von Selbstkontrolle bei einer Aufgabe sind Sie nicht dazu aufgelegt, sich bei einer weiteren erneut anzustrengen, obwohl Sie das tun könnten, wenn Sie es wirklich müssten. Bei mehreren Experimenten konnten die Probanden den Wirkungen einer Ego-Depletion widerstehen, wenn sie einen starken Anreiz dazu erhielten.[6] Dagegen ist vermehrte Anstrengung keine Option, wenn Sie sechs Ziffern im Kurzzeitgedächtnis halten müssen, während Sie eine Aufgabe erledigen. Ego-Depletion ist nicht der gleiche mentale Zustand wie kognitive Auslastung.

Die überraschendste Entdeckung, die Baumeisters Gruppe machte, zeigt, wie er es formuliert, dass die Idee einer mentalen Energie mehr als nur eine Metapher ist.[7] Das Nervensystem verbraucht mehr Glukose als die meisten anderen Körperteile, und anstrengende mentale Aktivität scheint in der Glukose-Währung besonders teuer zu sein. Wenn Sie intensiv über ein schwieriges Problem nachdenken oder eine Aufgabe ausführen, die Selbstkontrolle erfordert, sinkt Ihr Blutzuckerwert. Der Effekt ist ganz ähnlich wie bei einem Läufer, der beim Sprint die in seinen Muskeln gespeicherte Glukose aufbraucht. Daraus lässt sich die gewagte Schlussfolgerung ableiten, dass die Effekte der Ego-Depletion durch die Aufnahme von Glukose rückgängig gemacht werden könnten, und

Baumeister und seine Kollegen haben diese Hypothese in mehreren Experimenten bestätigt.[8]

Freiwillige in einer ihrer Studien sahen sich einen kurzen Stummfilm über eine Frau an, die interviewt wurde, und sie wurden gebeten, ihre Körpersprache zu interpretieren. Während sie die Aufgabe ausführten, wanderte eine Reihe von Wörtern in langsamer Folge über den Bildschirm. Die Teilnehmer wurden ausdrücklich angewiesen, die Wörter zu ignorieren, und sobald sie bemerkten, dass ihre Aufmerksamkeit abgelenkt wurde, sollten sie sich wieder auf das Verhalten der Frau konzentrieren. Man wusste, dass dieser Akt der Selbstkontrolle eine Ego-Depletion auslöst. Alle Freiwilligen tranken ein bisschen Limonade, bevor sie sich der zweiten Aufgabe zuwandten. Die Hälfte von ihnen trank mit Glukose gesüßte, die andere Hälfte mit Süßstoff versetzte Limonade. Anschließend bekamen alle Teilnehmer eine Aufgabe, bei der sie eine intuitive Reaktion überwinden mussten, um die richtige Antwort zu finden. Menschen mit erschöpfter Selbstregulation unterlaufen normalerweise viel mehr intuitive Fehler als nicht erschöpften Menschen, und diejenigen, die die mit Süßstoff versetzte Limonade tranken, zeigten den erwarteten Erschöpfungseffekt. Bei den Glukosetrinkern dagegen war die Selbstregulation nicht erschöpft. Die Wiederherstellung des üblichen Blutzuckerspiegels im Gehirn hatte verhindert, dass sich die Leistungsfähigkeit verschlechterte.

Es bedarf umfassender weiterer Forschungsarbeiten, um zu klären, ob die Aufgaben, die eine Glukose-Erschöpfung verursachen, auch das vorübergehende a*rousal* verursachen, das sich in der Zunahme der Pupillenweite und des Herzschlags widerspiegelt. Über eine verstörende Demonstration von Auswirkungen der Ego-Depletion auf Urteile wurde unlängst in der wissenschaftlichen Fachzeitschrift *Proceedings of the National Academy of Sciences* berichtet[9]. Die ahnungslosen Studienteilnehmer waren acht Bewährungsrichter in Israel. Sie verbrachten ganze Tage damit, Anträge auf bedingte Entlassung zu prüfen. Die Fälle wurden ihnen in zufälliger Reihenfolge vorgelegt, und die Richter verwendeten auf jeden einzelnen nur wenig Zeit, im Schnitt sechs Minuten. (Die Standardentscheidung ist die Ablehnung des Antrags auf bedingte Entlassung; nur 35 Prozent der Gesuche wurden positiv entschieden. Die genaue Zeit der Beschlussfassung wurde aufgezeichnet, und die Zeiten der drei Essenspausen der Richter – morgens, mittags und nachmittags –

wurden ebenfalls notiert.) Die Autoren der Studie trugen in einem Diagramm den Prozentsatz der bewilligten Anträge gegen die seit der letzten Essenspause vergangene Zeit auf. Der Prozentsatz gipfelte nach jedem Essen, wenn etwa 65 Prozent der Anträge bewilligt wurden. Im Verlauf der nächsten zwei Stunden, bis zur nächsten Speisung der Richter, sank die Bewilligungsquote stetig, auf etwa null unmittelbar vor dem nächsten Mahl. Sie ahnen vielleicht, dass dies ein unerwünschtes Ergebnis ist, und die Autoren überprüften sorgfältig viele alternative Erklärungen. Die bestmögliche Erklärung der Daten hält schlechte Neuigkeiten für uns bereit: Erschöpfte und hungrige Richter scheinen auf die leichtere Standardposition der Ablehnung von Bewährungsgesuchen zurückzufallen. Vermutlich spielen dabei sowohl Erschöpfung als auch Hunger eine Rolle.

## Das faule System 2

Eine der Hauptfunktionen von System 2 besteht darin, von System 1 »vorgeschlagene« Gedanken und Handlungen zu überwachen und zu kontrollieren; einigen davon erlaubt es, sich direkt im Verhalten auszudrücken, während es andere unterdrückt oder modifiziert.

Nehmen wir folgende einfache Denkaufgabe. Versuchen Sie nicht, sie zu lösen, sondern vertrauen Sie Ihrer Intuition:

> Ein Schläger und ein Ball kosten 1,10 Dollar.
> Der Schläger kostet einen Dollar mehr als der Ball.
> Wie viel kostet der Ball?

Ihnen fällt eine Zahl ein. Die Zahl ist selbstverständlich 10, nämlich 10 Cent. Die charakteristische Besonderheit dieser leichten Denkaufgabe besteht darin, dass sie eine Antwort nahelegt, die intuitiv verlockend und falsch ist. Berechnen Sie es, und Sie werden es sehen. Wenn der Ball 10 Cent kostet, dann betragen die Gesamtkosten 1,20 Dollar (10 Cent für den Ball und 1,10 Dollar für den Schläger), nicht 1,10 Dollar. Die richtige Antwort lautet 5 Cent. Mit Sicherheit fiel die intuitive Antwort auch denjenigen ein, die schließlich auf die richtige Zahl kamen – es gelang ihnen irgendwie, sich der Intuition zu widersetzen.

Shane Frederick und ich arbeiteten gemeinsam an einer Urteilstheorie auf der Basis zweier Systeme, und er benutzte die Schläger-und-Ball-Aufgabe, um eine zentrale Frage zu erkunden: Wie intensiv überwacht System 2 die Vorschläge von System 1? Dabei ging er davon aus, dass wir eine wichtige Tatsache über jeden wissen, der sagt, der Ball koste 10 Cent: Diese Person hat nicht aktiv die Richtigkeit der Antwort überprüft, und ihr System 2 unterstützte eine intuitive Lösung, die es mit einer geringfügigen Anstrengung hätte verwerfen können. Außerdem wissen wir, dass diejenigen, die die intuitive Antwort gegeben haben, einen offensichtlichen sozialen Hinweisreiz übersahen; sie hätten sich fragen müssen, wieso jemand eine Denkaufgabe mit einer so leichten Antwort in einen

Fragebogen aufnehmen würde. Es ist bemerkenswert, wenn die intuitive Antwort nicht überprüft wird, weil die Kosten des Überprüfens so gering sind: Einige wenige Sekunden mentaler Arbeit (das Problem ist mittelschwer) mit leicht angespannten Muskeln und vergrößerten Pupillen könnten einen peinlichen Fehler vermeiden. Menschen, die 10 Cent sagen, scheinen glühende Anhänger des Gesetzes des geringsten Aufwands zu sein. Personen, die diese Antwort vermeiden, scheinen einen aktiveren Intellekt zu besitzen.

Viele Tausend Studenten haben die Schläger-und-Ball-Denkaufgabe beantwortet, und die Ergebnisse sind erschreckend. Über 50 Prozent der Studenten an den Universitäten Harvard, MIT und Princeton gaben die intuitive – falsche – Antwort.[10] Bei Universitäten mit weniger strenger Auslese lag die Rate derjenigen, die nachweislich die intuitive Antwort nicht überprüften, bei über 80 Prozent. Das Schläger-Ball-Problem ist unsere erste Begegnung mit einer Beobachtung, die sich wie ein roter Faden durch dieses Buch zieht: Viele Menschen vertrauen ihren Intuitionen allzu sehr. Offensichtlich erleben sie kognitive Anstrengung zumindest als leicht unangenehm und meiden sie möglichst. Nun werde ich Ihnen ein logisches Argument vorstellen – zwei Prämissen und eine Schlussfolgerung. Versuchen Sie, so schnell wie möglich herauszufinden, ob das Argument gültig ist. Folgt die Schlussfolgerung aus den Prämissen?

> Alle Rosen sind Blumen.
> Einige Blumen verwelken schnell.
> Deshalb verwelken einige Rosen schnell.

Eine große Mehrheit der College-Studenten stuft diesen Syllogismus als gültig ein.[11] Tatsächlich ist das Argument nicht stichhaltig, weil es möglich ist, dass unter den rasch verwelkenden Blumen keine Rosen sind. Wie bei dem Schläger-und-Ball-Problem fällt einem sofort eine plausible Antwort ein. Es bedarf harter Arbeit, sich über sie hinwegzusetzen – die hartnäckige Vorstellung »Es ist wahr! Es ist wahr!« macht es schwer, die Logik zu überprüfen, und die meisten Menschen machen sich nicht die Mühe, das Problem zu durchdenken. Dieses Experiment lässt entmutigende

Rückschlüsse auf die Qualität des logischen Denkens im Alltagsleben zu. Es deutet darauf hin, dass Menschen, wenn sie eine Schlussfolgerung für wahr halten, höchstwahrscheinlich auch Argumente glauben, die diese Schlussfolgerung untermauern, auch wenn diese Argumente wenig stichhaltig sind. Wenn System 1 beteiligt ist, kommt die Schlussfolgerung zuerst, und die Argumente folgen.

Betrachten Sie als Nächstes die folgende Frage, und beantworten Sie sie schnell, bevor Sie weiterlesen.

> Wie viele Morde ereignen sich innerhalb eines Jahres im Bundesstaat Michigan?

Die Frage, die sich ebenfalls Shane Frederick ausgedacht hat, ist wieder eine Herausforderung für System 2. Hier kommt es darauf an, ob der Befragte sich daran erinnert, dass Detroit, eine Stadt mit hoher Kriminalität, in Michigan liegt. College-Studenten in den Vereinigten Staaten ist diese Tatsache bekannt, und sie werden Detroit richtigerweise als die größte Stadt in Michigan identifizieren. Aber die Kenntnis einer Tatsache ist nicht eine Frage von alles oder nichts. Fakten, die wir kennen, fallen uns nicht immer dann ein, wenn wir sie brauchen. Menschen, die sich daran erinnern, dass Detroit in Michigan liegt, schätzen die Mordrate in dem Bundesstaat höher ein als Menschen, die sich nicht daran erinnern. Aber eine Mehrheit derjenigen, die von Frederick befragt wurden, dachte nicht an die Stadt, als man sie nach dem Bundesstaat fragte. Tatsächlich geben Personen, die nach Michigan gefragt wurden, im Schnitt *niedrigere* Schätzungen ab als Mitglieder einer ähnlichen Gruppe, die nach der Mordrate in Detroit gefragt wurden.

Schuld an der Tatsache, dass bestimmten Personen Detroit nicht einfiel, haben sowohl System 1 als auch System 2. Ob jemandem die Stadt einfällt, wenn der Bundesstaat erwähnt wird, hängt zum Teil von der automatischen Funktionsweise des Gedächtnisses ab. Menschen unterscheiden sich in dieser Hinsicht. Der Bundesstaat Michigan ist im Gehirn mancher Menschen sehr detailliert repräsentiert: Menschen, die in dem Bundesstaat leben, erinnern sich eher an viele Fakten, die ihn betreffen, als Menschen, die irgendwo anders leben; Geografie-Fans erinnern sich an mehr als

andere, die sich auf Baseball-Statistik spezialisiert haben; intelligentere Personen haben eher als andere differenziertere Repräsentationen von den meisten Dingen. Intelligenz ist nicht nur die Fähigkeit zu logischem Denken; sie ist auch die Fähigkeit, im Gedächtnis relevantes Material aufzufinden und die Aufmerksamkeit bei Bedarf effektiv zu nutzen. Die Gedächtnisfunktion ist ein Attribut von System 1. Doch jeder hat die Option, einen Gang zurückzuschalten, um eine aktive Gedächtnissuche nach allen möglicherweise relevanten Fakten durchzuführen – so wie jeder auf die Bremse hätte treten können, um die intuitive Antwort bei dem Schläger-und-Ball-Problem zu überprüfen. Das Ausmaß gezielter Überprüfung und Suche ist ein Kennzeichen von System 2 und unterscheidet sich von Mensch zu Mensch.

Das Schläger-und-Ball-Problem, der Blumen-Syllogismus und das Michigan-Detroit-Problem haben etwas gemeinsam. Wenn man an diesen Minitests scheitert, scheint dies, zumindest bis zu einem gewissen Grad, auf unzureichende Motivation zurückzuführen zu sein, darauf, dass man sich nicht genug Mühe gibt. Jeder, der zum Studium an einer angesehenen Universität zugelassen wird, ist zweifellos in der Lage, die ersten beiden Fragen logisch klar zu durchdenken und über Michigan so lange nachzudenken, dass er sich an die größte Stadt in diesem Bundesstaat und ihre hohe Kriminalität erinnert. Diese Studenten können viel schwierigere Probleme lösen, wenn sie sich nicht mit einer oberflächlich plausiblen Antwort, die ihnen spontan einfällt, zufriedengeben. Es ist jedoch verstörend, wenn man sieht, wie leicht sie zufriedenzustellen sind und dann nicht weiter nachdenken. »Faul« ist ein hartes Urteil über die mentale Selbstüberwachung dieser jungen Menschen und ihr System 2, aber es scheint nicht unangemessen zu sein. Diejenigen, die die Sünde der intellektuellen Trägheit vermeiden, könnten »engagiert« genannt werden. Sie sind aufgeweckter, intellektuell aktiver, weniger gewillt, sich mit oberflächlich attraktiven Antworten zu begnügen, und kritischer gegenüber ihren Intuitionen. Der Psychologe Keith Stanovich würde sie als rationaler bezeichnen. [12]

## Intelligenz, Kontrolle und Rationalität

Forscher haben unterschiedliche Methoden angewandt, um den Zusammenhang zwischen Denken und Selbstkontrolle zu erforschen. Einige sind dabei so vorgegangen, dass sie nach der Korrelation gefragt haben: Wenn Menschen nach ihrer Selbstkontrolle und nach ihren kognitiven Fähigkeiten eingestuft würden, hätten sie dann in den beiden Rankings ähnliche Rangplätze?

In einem der berühmtesten Experimente in der Geschichte der Psychologie konfrontierten Walter Mischel und seine Studenten vierjährige Kinder mit einem grausamen Dilemma.[13] Sie hatten die Wahl zwischen einer kleinen Belohnung (einem Keks), die sie jederzeit bekommen konnten, und einer größeren (zwei Keksen), für die sie 15 Minuten unter schwierigen Bedingungen warten mussten. Sie hatten in einem Raum allein an einem Schreibtisch zu sitzen, auf dem sich zwei Objekte befanden: ein Keks und eine Glocke, die das Kind jederzeit läuten konnte, um den Experimentator hereinzurufen, damit er ihm den Keks gibt. Das Experiment wurde folgendermaßen beschrieben: »Es gab kein Spielzeug, keine Bücher, keine Bilder oder andere, potenziell ablenkende Gegenstände in dem Raum. Der Experimentator verließ den Raum und kehrte erst nach 15 Minuten zurück, beziehungsweise dann, wenn das Kind die Glocke geläutet hatte, die Belohnungen verzehrt hatte, aufgestanden war oder Anzeichen von Stress gezeigt hatte.«[14]

Die Kinder wurden durch einen Einwegspiegel beobachtet, und der Film, der ihr Verhalten während der Wartezeit zeigt, ließ die Zuschauer immer in schallendes Gelächter ausbrechen. Etwa der Hälfte der Kinder gelang die Meisterleistung, 15 Minuten zu warten, hauptsächlich dadurch, dass sie ihre Aufmerksamkeit von der verlockenden Belohnung ablenkten. 10, 15 Jahre später hatte sich eine große Kluft aufgetan zwischen denjenigen, die der Belohnung widerstanden hatten, und denjenigen, die ihr erlegen waren. Diejenigen, die widerstehen konnten, hatten höhere Messwerte exekutiver Kontrolle bei kognitiven Aufgaben, und sie konnten insbesondere ihre Aufmerksamkeit effektiver steuern. Als junge Erwachsene waren sie wenig anfällig für Drogenmissbrauch. Bei den intellektuellen Fähigkeiten zeigte

sich ein deutlicher Unterschied: Die Kinder, die als Vierjährige mehr Selbstkontrolle gezeigt hatten, erzielten bei Intelligenztests erheblich höhere IQ-Werte.[15]

Eine Forschergruppe an der Universität Oregon studierte mit verschiedenen Methoden den Zusammenhang zwischen kognitiver Kontrolle und Intelligenz; so versuchten sie unter anderem auch, die Intelligenz durch Verbesserung der Aufmerksamkeitssteuerung zu steigern. In fünf vierzigminütigen Sitzungen ließen sie Kinder zwischen vier und sechs Jahren verschiedene Computerspiele spielen, die speziell Aufmerksamkeit und kognitive Kontrolle beanspruchten. Bei einer der Übungen sollten die Kinder mit einem Joystick eine Zeichentrickkatze verfolgen und sie unter Umgehung eines Schlammpfuhls auf eine Wiese führen. Nun schrumpfte die Grasfläche nach und nach, während sich das schlammige Areal ausweitete, sodass der Joystick immer präziser gesteuert werden musste. Die Experimentatoren fanden heraus, dass die Schulung der Aufmerksamkeit nicht nur die exekutive Kontrolle verbesserte; auch die Leistung in Tests der nonverbalen Intelligenz erhöhte sich, und die Leistungssteigerung hielt mehrere Monate an.[16] Im Rahmen weiterer Studien identifizierte dieselbe Gruppe spezifische Gene, die an der Aufmerksamkeitssteuerung beteiligt sind. Sie zeigte, dass elterliches Erziehungsverhalten sich ebenfalls auf diese Fähigkeit auswirkte, und wies nach, dass ein enger Zusammenhang zwischen der Fähigkeit der Kinder, ihre Aufmerksamkeit zu steuern, und ihrer Fähigkeit, ihre Emotionen zu kontrollieren, besteht.

Shane Frederick entwickelte den sogenannten *Cognitive Reflection Test* (CRT), der aus dem Schläger-und-Ball-Problem und zwei weiteren Fragen besteht, die er deshalb ausgewählt hat, weil sie ebenfalls eine intuitive Antwort herausfordern, die sowohl bestechend als auch falsch ist (die Fragen werden in Kapitel 5 vorgestellt). Anschließend untersuchte er die Merkmale der Studenten, die bei diesem Test sehr schlecht abschnitten – bei ihnen ist die Überwachungsfunktion von System 2 sehr schwach –, und er stellte fest, dass sie dazu neigen, Fragen mit der ersten Idee zu beantworten, die ihnen einfällt, und dass sie nicht gewillt sind, die Anstrengung aufzubringen, die notwendig wäre, um ihre Intuitionen zu überprüfen. Personen, die bei Denkaufgaben unkritisch ihren Intuitionen folgen, neigen auch dazu, andere Vorschläge von System 1 unhinterfragt anzunehmen. So

sind sie insbesondere impulsiv, ungeduldig und erpicht auf sofortige Belohnung. So sagen zum Beispiel 63 Prozent derjenigen, die intuitiv antworten, dass ihnen 3400 Dollar in diesem Monat lieber wären als 3800 Dollar im nächsten. Nur 37 Prozent derjenigen, die alle drei Denkaufgaben richtig beantworten, haben die gleiche kurzsichtige Präferenz dafür, sofort einen kleineren Betrag zu erhalten. Danach gefragt, wie viel mehr sie für eine Expresszustellung des von ihnen bestellten Buches zu zahlen bereit sind, sind diejenigen, die beim *Cognitive Reflection Test* niedrige Punktwerte erzielen, gewillt, doppelt so viel zu bezahlen wie diejenigen, die dabei gut abschneiden. Fredericks Befunde deuten darauf hin, dass die Figuren unseres Psychodramas verschiedene »Persönlichkeiten« haben. System 1 ist impulsiv und intuitiv; System 2 kann logisch denken, und es ist bedächtig, aber zumindest bei einigen Menschen ist es auch faul. Wir erkennen ähnliche Unterschiede zwischen Menschen: Einige Personen gleichen mehr ihrem System 2; andere stehen ihrem System 1 näher. Es hat sich gezeigt, dass dieser einfache Test Denkfaulheit relativ zuverlässig vorhersagen kann.

Keith Stanovich und sein langjähriger Mitarbeiter Richard West führten ursprünglich die Begriffe System 1 und System 2 ein (heute ziehen sie es vor, von Typ-1- und Typ-2-Prozessen zu sprechen). Stanovich und seine Mitarbeiter haben jahrzehntelang die Unterschiede zwischen Individuen bei der Lösung jener Art von Problemen erforscht, mit denen sich dieses Buch befasst. Sie haben eine grundlegende Frage in vielfältiger Variation gestellt: Wieso sind einige Menschen anfälliger für Urteilsfehler als andere? Stanovich veröffentlichte seine Schlussfolgerungen in einem Buch mit dem Titel *Rationality and the Reflective Mind*, in dem er auf eine kühne und unverwechselbare Weise an das Thema dieses Kapitels herangeht. Er unterscheidet klar und deutlich zwischen zwei Teilen von System 2 – seine Unterscheidung geht so weit, dass er von zwei verschiedenen »Denkvermögen« *(minds)* spricht. Eines dieser kognitiven Module (er nennt es »algorithmisch«) befasst sich mit langsamem Denken und anspruchsvoller Berechnung. Einige Menschen kommen mit diesen Intelligenzaufgaben besser zurecht als andere – sie erzielen bei Intelligenztests sehr hohe Punktwerte, und sie können rasch und effizient von einer Aufgabe zu einer anderen wechseln. Stanovich behauptet nun jedoch, dass hohe Intelligenz Menschen nicht immun gegen kognitive Verzerrungen macht. Hier kommt eine weitere Fähigkeit ins Spiel, die er

»Rationalität« nennt. Stanovichs Konzept einer rationalen Person deckt sich weitgehend mit dem, was ich weiter vorn »[mental] engagiert« genannt habe. Der Kern seines Arguments lautet, dass *Rationalität* von *Intelligenz* unterschieden werden sollte. Seiner Auffassung nach ist oberflächliches oder »faules« Denken eine Unzulänglichkeit des reflexiven Intellekts, ein Versagen der Rationalität. Dies ist eine verlockende Idee, die zum Nachdenken anregt. Stanovich und seine Mitarbeiter haben etwas herausgefunden, was ihre Annahme untermauert: Die Schläger-und-Ball-Frage und andere, ähnliche Fragen sind etwas bessere Indikatoren unserer Anfälligkeit für kognitive Fehler als herkömmliche Intelligenzmaße wie IQ-Tests.[17] Die Zeit wird erweisen, ob die Unterscheidung zwischen Intelligenz und Rationalität zu neuen Entdeckungen führen kann.

## Zum Thema »Kontrolle«

»Es kostete sie keine Mühe, sich stundenlang auf eine Aufgabe zu konzentrieren. Sie befand sich in einem Zustand des Flow.«

»Sein Ego war nach einem langen Arbeitstag voller Sitzungen erschöpft. Daher schaltete er auf Standardvorgehensweise, statt das Problem zu durchdenken.«

»Er machte sich nicht die Mühe, zu überprüfen, ob das, was er sagte, Hand und Fuß hatte. Hat er grundsätzlich ein faules System 2, oder war er ungewöhnlich erschöpft?«

»Leider spricht sie in der Regel immer das aus, was ihr als Erstes in den Sinn kommt.Vermutlich fällt es ihr auch schwer, Belohnungen aufzuschieben. Schwaches System 2.«

# *4. Die Assoziationsmaschine* [1]

Lassen Sie uns unsere Erkundung der überraschenden Funktionsmechanismen von System 1 mit den folgenden Wörtern beginnen:

Bananen  Erbrechen

In den letzten ein, zwei Sekunden ist viel in Ihnen abgelaufen. Sie haben einige unangenehme Bilder und Erinnerungen erlebt. Sie haben Ihr Gesicht ein wenig verzogen, in einem Ausdruck des Ekels, und Sie haben dieses Buch vielleicht unmerklich ein Stückchen von sich weggeschoben. Ihr Herz schlägt schneller, die Haare an Ihren Armen haben sich leicht aufgerichtet, und Ihre Schweißdrüsen wurden aktiviert. Kurzum, Sie reagierten auf das ekelerregenden Wort mit einer abgeschwächten Version dessen, wie Sie auf das reale Ereignis reagieren würden. All dies geschah vollkommen unwillkürlich, jenseits Ihrer Kontrolle.[2]

Ohne einen besonderen Grund dafür zu haben, ging Ihr Intellekt automatisch von einer zeitlichen Abfolge und einem kausalen Zusammenhang zwischen den Wörtern »Bananen« und »Erbrechen« aus und erstellte ein vages Szenario, in dem Bananen die Ursache der Übelkeit waren. Infolgedessen empfinden Sie eine vorübergehende Abneigung gegen Bananen (keine Sorge, sie ist nicht von Dauer). Der Zustand Ihres Gedächtnisses hat sich auch noch in anderer Weise verändert: Sie sind jetzt in einer erhöhten Bereitschaft, Objekte und Konzepte zu erkennen – und darauf zu reagieren –, die mit »Erbrechen« assoziiert sind, wie etwa »Krankheit«, »Gestank« und »Übelkeit«, und Wörter, die mit »Bananen« verbunden sind, wie »gelb« und »Obst« und vielleicht »Äpfel« und »Beeren«.

Erbrechen ereignet sich normalerweise in bestimmten Situationen, etwa wenn man einen Kater hat oder an Magenverstimmung leidet. Sie befinden sich jetzt auch in einer erhöhten Bereitschaft, Wörter zu erkennen, die mit anderen Ursachen des gleichen unerfreulichen Resultats assoziiert sind. Außerdem bemerkte Ihr System 1 die Tatsache, dass das Nebeneinander der beiden Wörter ungewöhnlich ist; es ist Ihnen vermutlich vorher noch nie begegnet. Sie waren leicht überrascht.

Diese komplexe Konstellation von Reaktionen ereignete sich schnell, automatisch und mühelos. Sie haben sie nicht willentlich herbeigeführt, und Sie konnten sie nicht aufhalten. Es war eine Operation von System 1. Den Ereignissen, die dadurch ausgelöst wurden, dass Sie die Wörter sahen, liegt ein Prozess zugrunde, der »assoziative Aktivierung« genannt wird: Vorstellungen, die wachgerufen wurden, lösen viele andere Vorstellungen aus, in einer sich ausbreitenden Aktivitätskaskade in Ihrem Gehirn. Das Kennzeichen dieser komplexen Menge mentaler Ereignisse ist ihre Kohärenz. Jedes Element ist mit anderen verbunden, und jedes stützt und stärkt die anderen. Ein Wort ruft Erinnerungen wach, die Emotionen wachrufen, die ihrerseits Gesichtsausdrücke und andere Reaktionen hervorrufen, wie etwa eine allgemeine Anspannung und eine Tendenz zur Vermeidung. Der Gesichtsausdruck und die Bewegung der Abwendung verstärken die Gefühle, mit denen sie verbunden sind, und die Gefühle ihrerseits verstärken kompatible Vorstellungen. All dies geschieht schnell und gleichzeitig und erzeugt ein sich selbst verstärkendes Muster kognitiver, emotionaler und körperlicher Reaktionen, das sowohl vielfältig als auch integriert ist und »assoziativ kohärent« genannt wird.

Innerhalb von etwa einer Sekunde gelang Ihnen, automatisch und unbewusst, ein bemerkenswertes Kunststück. Von einem völlig unerwarteten Ereignis ausgehend, versuchte Ihr System 1, die Situation so gründlich wie nur möglich zu verstehen – zwei einfache Wörter, in ungewohnter Weise nebeneinanderstehend –, indem es die Wörter in einer kausalen Geschichte miteinander verknüpfte; es schätzte die mögliche Bedrohung ein (gering bis moderat) und schuf einen Kontext für zukünftige Entwicklungen, indem es Sie auf Ereignisse vorbereitete, die gerade wahrscheinlicher geworden waren; es erzeugte auch einen Kontext für das gegenwärtige Ereignis, indem es beurteilte, wie überraschend es war. Sie wussten schließlich so gut über die Vergangenheit Bescheid und waren so gut für die Zukunft gerüstet, wie es überhaupt möglich war.

Ihr System 2 hat dabei seltsamerweise die bloße Verknüpfung von zwei Wörtern als eine Repräsentation der Wirklichkeit behandelt. Ihr Körper reagierte in abgeschwächter Weise so, wie er auf das reale Ereignis reagieren würde, und die emotionale Reaktion wie auch das körperliche Zurückweichen waren Teil der Interpretation des Ereignisses. Kognitionswissenschaftler haben in den letzten Jahren betont, dass Kognition leiblich verankert ist; wir denken mit unserem Körper, nicht nur mit unserem Gehirn.[3]

Der Mechanismus, der diese mentalen Ereignisse verursacht, ist seit langer Zeit bekannt: Es ist die Assoziation von Vorstellungen. Wir alle wissen aus Erfahrung, dass Vorstellungen in unserem Bewusstsein in recht geordneter Weise aufeinanderfolgen. Die britischen Philosophen des 17. und 18. Jahrhunderts suchten nach den Regeln, die solche Folgen erklären. In der 1748 veröffentlichen Abhandlung *Eine Untersuchung über den menschlichen Verstand (An Enquiry Concerning Human Understanding)* unterschied der schottische Philosoph David Hume drei Grundprinzipien der Assoziation: Ähnlichkeit, Kontiguität in Zeit und Raum und Kausalität. Unser Assoziationsbegriff hat sich seit Humes Zeit grundlegend gewandelt, aber seine drei Prinzipien bilden noch immer einen guten Ausgangspunkt.

Ich werde hier einen sehr weit gefassten Vorstellungsbegriff verwenden. Eine Vorstellung kann konkret oder abstrakt sein, und sie kann in vielfältiger Weise zum Ausdruck gebracht werden: durch ein Verb, ein Substantiv, ein Adjektiv oder eine geballte Faust. Für Psychologen sind Vorstellungen Knoten in einem riesigen Netzwerk, assoziatives Gedächtnis genannt, in dem jede Vorstellung mit vielen anderen verbunden ist. Es gibt unterschiedliche Typen von Verknüpfungen: Ursachen sind mit ihren Wirkungen verknüpft (Virus → Erkältung); Objekte mit ihren Eigenschaften (Linde → grün); Objekte mit den Kategorien, zu denen sie gehören (Banane → Frucht). Ein Fortschritt gegenüber Hume liegt darin, dass wir heute nicht mehr annehmen, dass das Denken aus einer Folge bewusster Vorstellungen besteht, die sich wie Perlen an einer Kette aneinanderreihen. Nach der neuesten Theorie über die Funktionsweise des assoziativen Gedächtnisses laufen viele Prozesse gleichzeitig ab. Eine Vorstellung, die aktiviert wurde, ruft nicht einfach nur eine andere Vorstellung wach. Sie aktiviert viele Vorstellungen, die ihrerseits weitere Vorstellungen evozieren. Außerdem werden nur einige wenige der aktivierten Vorstellungen bewusst registriert; der größte Teil der Arbeit des

assoziativen Denkens vollzieht sich lautlos, unterhalb der Bewusstseinsschwelle. Die Auffassung, dass wir nur beschränkten Zugang zum »Innenleben« unseres Geistes haben, ist schwer zu akzeptieren, weil es nicht unserem eigenen Erleben entspricht, aber sie ist trotzdem wahr: Wir wissen viel weniger über uns selbst, als wir zu wissen glauben.

## Die Wunder des Priming

Wie in der Wissenschaft üblich, ist der erste große Durchbruch in unserem Verständnis des Assoziationsmechanismus der Verbesserung einer Messmethode zu verdanken. Bis vor ein paar Jahrzehnten konnte man Assoziationen nur dadurch erforschen, dass man Menschen Fragen stellte wie: »Was ist das erste Wort, an das Sie denken, wenn Sie das Wort ›Tag‹ hören?« Die Forscher zählten die Häufigkeiten von Antworten wie »Nacht«, »sonnig« oder »lang«. In den 1980er-Jahren fanden Psychologen dann heraus, dass die Darbietung eines Wortes sofortige messbare Veränderungen in der Leichtigkeit verursacht, mit der verwandte Wörter ins Gedächtnis gerufen werden können. Wenn Sie vor Kurzem das Wort *eat* (»Essen«) gesehen oder gehört haben, werden Sie vorübergehend das Wortfragment *so_p* eher als *soup* (»Suppe«) denn als *soap* (»Seife«) vervollständigen. Das Umgekehrte wäre der Fall, wenn sie gerade *wash* (»Waschen«) gesehen hätten. Wir nennen dies einen »Priming-Effekt« (»Bahnungs-Effekt«) und sprechen davon, dass die Vorstellung »essen« die Vorstellung »Suppe« bahnt, während »waschen« die Vorstellung »Seife« bahnt.[4] Priming-Effekte können viele verschiedene Formen annehmen. Wenn Ihnen gerade die Vorstellung »essen« durch den Kopf geht (unabhängig davon, ob sie Ihnen bewusst ist oder nicht), erkennen Sie das Wort »Suppe« schneller als sonst, wenn es geflüstert wird oder in einer verschwommenen Schrift erscheint. Und natürlich sind Sie nicht nur für die Vorstellung »Suppe« geprimt, sondern auch für eine Vielzahl weiterer Vorstellungen rund ums Essen, wie etwa »Gabel«, »hungrig«, »Fett«, »Diät« und »Keks«. Wenn Sie bei Ihrer letzten Mahlzeit an einem wackligen Restauranttisch saßen, werden Sie bei der kognitiven Informationsverarbeitung auch »wacklig« mit höherer Wahrscheinlichkeit aus Ihrem Gedächtnis abrufen. Außerdem können die geprimten Vorstellungen ihrerseits in gewissem Umfang andere Vorstellungen primen, wenn auch schwächer. Wie Kräuselwellen auf der Oberfläche eines Teichs breitet sich die Aktivierung durch einen kleinen Teil des riesigen Netzwerks assoziierter Vorstellungen aus. Die Kartierung dieser Kräuselwellen ist heute eines der spannendsten psychologischen Forschungsfelder.

Ein weiterer großer Fortschritt im Verständnis der Arbeitsweise des Gedächtnisses war die Entdeckung, dass Priming nicht auf Konzepte und Wörter beschränkt ist. Auch wenn sich dies nicht mit unserem bewussten Erleben deckt, müssen wir uns mit der befremdlichen Vorstellung abfinden, dass unsere Handlungen und Emotionen durch Ereignisse geprimt werden können, deren wir uns nicht einmal bewusst sind. In einem Experiment, das auf Anhieb zu einem Klassiker wurde, baten der Psychologe John Bargh und seine Mitarbeiter Studenten der Universität von New York – die größtenteils zwischen 18 und 22 Jahre alt waren –, aus einer Menge von fünf Wörtern (zum Beispiel »findet er es gelb sofort«) Vier-Wort-Sätze zu bilden.[5] Bei einer Gruppe von Studenten enthielt die Hälfte der ungeordneten Sätze Wörter, die – in den Vereinigten Staaten – mit älteren Menschen assoziiert werden, wie »Florida«, »vergesslich«, »glatzköpfig«, »grau« oder »Falte«. Als sie diese Aufgabe beendet hatten, wurden die jungen Versuchsteilnehmer für ein weiteres Experiment in ein Büro geschickt, das am Ende desselben Flurs lag. Dieser kurze Spaziergang war der entscheidende Punkt in diesem Experiment. Die Forscher maßen unauffällig die Zeit, die die Probanden benötigten, um von einem Ende des Flurs ans andere zu gelangen. Wie von Bargh erwartet, gingen die jungen Leute, die einen Satz aus »altersbezogenen« Wörtern gebildet hatten, erheblich langsamer durch den Flur als die anderen.

Der »Florida-Effekt« umfasst zwei Priming-Phasen. Zunächst primt die Menge der Wörter Gedanken an hohes Alter, obwohl das Wort »alt« nie erwähnt wird; anschließend primen diese Gedanken ein Verhalten, langsames Gehen, das mit Betagtheit assoziiert ist. All dies geschieht unbewusst. Als die Studenten im Anschluss gefragt wurden, sagten sie alle, sie hätten nicht bemerkt, dass die Wörter ein gemeinsames Thema haben, und sie alle beteuerten, dass nichts von dem, was sie nach dem ersten Experiment taten, von den Wörtern, die ihnen dargeboten worden waren, beeinflusst worden sei. Die Vorstellung »Alter« war ihnen nicht bewusst geworden, und doch hatte sich ihr Verhalten verändert. Dieses bemerkenswerte Priming-Phänomen – die Beeinflussung einer Handlung durch eine Vorstellung – wird ideomotorischer Effekt genannt. Obwohl Sie es bestimmt nicht bemerkt haben, hat die Lektüre dieses Absatzes Sie ebenfalls geprimt. Wenn Sie aufgestanden sein sollten, um sich ein Glas Wasser zu holen, dürften Sie sich etwas langsamer von Ihrem Stuhl erhoben haben als üblich – es sei denn, Sie mögen keine älteren Menschen; in

diesem Fall deuten die Forschungsergebnisse darauf hin, dass Sie ein wenig schneller aufgestanden sein dürften!

Die ideomotorische Verknüpfung funktioniert auch andersherum. Eine Studie, die an einer deutschen Universität durchgeführt wurde, war das Spiegelbild des früheren Experiments, das Bargh und seine Mitarbeiter in New York durchgeführt hatten. Studenten wurden gebeten, fünf Minuten lang mit einer Geschwindigkeit von dreißig Schritten pro Minute, was etwa ein Drittel ihres normalen Tempos war, in einem Raum umherzugehen. Nach dieser kurzen Übung erkannten die Teilnehmer Wörter, die sich auf hohes Alter bezogen, wie etwa »vergesslich«, »alt« und »einsam«, sehr viel schneller wieder.[6] Wechselseitige Priming-Effekte erzeugen eine kohärente Reaktion: Wenn man geprimt wurde, an hohes Alter zu denken, neigt man zu seniorentypischem Verhalten, und wenn man sich wie ein älterer Mensch verhält, verstärkt dies den Gedanken an hohes Alter.

Reziproke Verknüpfungen sind in dem assoziativen Netzwerk weit verbreitet. So lächeln Sie im Allgemeinen, wenn Sie vergnügt sind, und wenn Sie lächeln, erheitert Sie das. Nehmen Sie einen Bleistift, und klemmen Sie ihn ein paar Sekunden lang so zwischen Ihre Zähne, dass der Radiergummi nach rechts und die Spitze nach links zeigt. Halten Sie den Bleistift jetzt so, dass die Spitze geradeaus zeigt, indem Sie Ihre Lippen um das Ende mit dem Radiergummi schürzen. Sie haben vermutlich nicht bemerkt, dass eine dieser Handlungen Sie dazu zwang, die Stirn zu runzeln, während Sie bei der anderen das Gesicht unwillkürlich zu einem Lächeln verzogen. College-Studenten sollten beurteilen, wie witzig Cartoons aus Gary Larsons Reihe *The Far Side* sind, während sie einen Bleistift im Mund hielten.[7] Diejenigen, die »lächelten« (ohne sich dessen bewusst zu sein), fanden die Cartoons lustiger als diejenigen, die »die Stirn runzelten«. In einem anderen Experiment berichteten Personen, die die Stirn runzelten (indem sie die Augenbrauen zusammenzogen), von einer verstärkten emotionalen Reaktion auf erschütternde Bilder von hungernden Kindern, streitenden Menschen und verstümmelten Unfallopfer.[8]

Einfache, weitverbreitete Gesten können ebenfalls unbewusst unsere Gedanken und Gefühle beeinflussen. In einem Experiment sollten die Versuchspersonen durch neue Kopfhörer Botschaften lauschen.[9] Ihnen wurde gesagt, Zweck des Experiments sei es, die Qualität der Audiogeräte zu testen, und sie sollten ihre Köpfe wiederholt bewegen, um mögliche

Klangverzerrungen festzustellen. Die Hälfte der Teilnehmer sollte mit dem Kopf nicken, während die anderen den Kopf schütteln sollten. Die Nachrichten, die ihnen vorgespielt wurden, waren Radiokommentare. Diejenigen, die nickten (eine bejahende Geste), stimmten der Nachricht, die Sie hörten, im Allgemeinen zu, während diejenigen, die den Kopf schüttelten, sie tendenziell ablehnten. Wieder waren sich die Probanden dessen nicht bewusst, vielmehr bestand nur eine gewohnheitsmäßige Beziehung zwischen einer ablehnenden oder zustimmenden Einstellung und ihrem üblichen mimischen Ausdruck. Sie verstehen jetzt, wieso die geläufige Ermahnung, »sich ruhig und freundlich zu verhalten, egal wie man sich fühlt«, ein sehr guter Ratschlag ist: Wenn Sie ihn befolgen, werden Sie wahrscheinlich damit belohnt, dass Sie sich ruhig und gelassen fühlen.

## Primes, die uns anleiten

Studien über Priming-Effekte führten zu Entdeckungen, die unser Selbstbild, bewusste und autonome Urheber unserer Urteile und Entscheidungen zu sein, infrage stellen. So glauben zum Beispiel die meisten von uns, unser Abstimmungsverhalten bei Wahlen sei ein bewusster Willensakt, in dem sich unsere Wertvorstellungen und unsere Urteile über politisches Handeln widerspiegelten und der nicht von Nebensächlichkeiten beeinflusst werde. So sollte unser Abstimmungsverhalten zum Beispiel nicht vom Standort des Wahllokals beeinflusst werden – was jedoch der Fall ist. Eine Studie über das Abstimmungsmuster in Wahlbezirken von Arizona im Jahr 2000 zeigte, dass die Unterstützung für Vorschläge zur Erhöhung der Bildungsausgaben erheblich größer war, wenn sich das Wahllokal in einer Schule befand statt nur in der Nähe einer Schule.[10] Ein anderes Experiment zeigte, dass die Bereitschaft von Personen, eine Schulinitiative zu unterstützen, auch dadurch gesteigert wurde, dass man ihnen Bilder von Klassenzimmern und Schulspinden zeigte. Der Effekt der Bilder war größer als der Unterschied zwischen Eltern und anderen Wählern! Die Erforschung des Priming-Phänomens ist ein gutes Stück über die anfänglichen Experimente, die zeigten, dass Menschen langsamer gehen, wenn man sie ans Alter erinnert, hinausgelangt. Wir wissen heute, dass Priming-Effekte in jeden Bereich unseres Lebens hineinreichen.

Ruft man das Thema Geld ins Gedächtnis, erzeugt dies einige sehr verstörende Effekte.[11] Den Teilnehmern eines Experiments wurde eine Liste mit fünf Wörtern gezeigt, aus denen Sie einen Vier-Wörter-Ausdruck mit einem Geld-Thema konstruieren sollten (»hoch ein bezahlten Job Schreibtisch« wurde »ein hoch bezahlter Job«). Andere Primes waren viel subtiler, etwa die Anwesenheit eines irrelevanten geldbezogenen Objekts im Hintergrund wie eines Stapels Monopoly-Geld auf einem Tisch oder eines Computers mit einem Bildschirmschoner von Dollarscheinen, die im Wasser treiben.

Auf Geld geprimte Personen werden unabhängiger, als sie es ohne den assoziativen Auslöser wären. Sie hielten bei dem Versuch, ein sehr schwieriges Problem zu lösen, fast doppelt so lange durch, ehe sie den

Experimentator um Hilfe baten – eine klare Demonstration erhöhter Selbstständigkeit. Auf Geld geprimte Menschen sind auch egoistischer: Sie waren weit weniger gewillt, einem anderen Studenten zu helfen, der vorgab, eine Experimentalaufgabe nicht zu verstehen. Als ein Experimentator unbeholfen ein Bündel Bleistifte auf den Boden fallen ließ, hoben die Teilnehmer, die (unbewusst) an Geld dachten, weniger Bleistifte auf. Bei einem anderen Experiment dieser Serie wurde den Probanden mitgeteilt, sie würden in Kürze ein Gespräch zum Kennenlernen mit einer anderen Person führen, und sie wurden gebeten, zu diesem Zweck zwei Stühle hinzustellen, während der Experimentator den Raum verließ, um diese Person holen zu gehen. Die auf Geld geprimten Teilnehmer stellten die Stühle viel weiter auseinander als ihre nicht geprimten Kollegen (118 gegenüber 80 Zentimeter). Auf Geld geprimte Studenten zeigten auch eine größere Präferenz fürs Alleinsein.

Der gemeinsame Nenner dieser Befunde besteht darin, dass die Idee des Geldes den Individualismus stärkt: das Widerstreben, sich mit anderen einzulassen, von anderen abhängig zu sein oder Forderungen von anderen anzunehmen. Die Psychologin, die diese bemerkenswerten Forschungen durchgeführt hat, Kathleen Vohs, hielt sich in vorbildlicher Weise mit Aussagen über die Bedeutung ihrer Befunde zurück und überließ diese Aufgabe ihren Lesern. Ihre Experimente sind tiefgründig – ihre Ergebnisse deuten darauf hin, dass das Leben in einer Kultur, die uns ständig an das Thema Geld erinnert, unser Verhalten und unsere Einstellungen womöglich in Weisen beeinflusst, von denen wir nichts wissen und auf die wir nicht stolz sein mögen. Einige Kulturen erinnern ihre Mitglieder häufig an Respekt, andere erinnern ihre Mitglieder an Gott, und einige Gesellschaften stärken durch große Bilder des »Lieben Führers« den Gehorsam. Kann ein Zweifel daran bestehen, dass die allgegenwärtigen Porträts des nationalen Führers in diktatorischen Gesellschaften nicht nur das Gefühl »Big Brother is Watching« vermitteln, sondern auch spontanes Denken und eigenständiges Handeln hemmen?

Die Ergebnisse von Priming-Studien legen die Vermutung nahe, dass die Anziehungskraft autoritärer Ideen wächst, wenn Menschen an ihre Sterblichkeit erinnert werden, da diese Ideen angesichts des Schreckens, der vom Tod ausgeht, beruhigend wirken mögen.[12] Andere Experimente haben freudsche Erkenntnisse über die Rolle von Symbolen und Metaphern bei unbewussten Assoziationen bestätigt. Nehmen wir zum Beispiel die

zweideutigen englischen Wortfragmente w _ _ h und s _ _ p. Personen, die in letzter Zeit aufgefordert wurden, an eine Handlung zu denken, für die sie sich schämten, vervollständigen diese Fragmente mit höherer Wahrscheinlichkeit als *wash* und *soup* und mit geringerer Wahrscheinlichkeit als *wish* (»Wunsch«) und *soup*. Außerdem veranlasst schon der bloße Gedanke daran, einem Arbeitskollegen in den Rücken zu fallen, Menschen dazu, eher Seife, Desinfektionsmittel oder Spülmittel als Batterien, Saft oder Zuckerstangen zu kaufen. Das Gefühl, sich psychisch beschmutzt zu haben, scheint den Wunsch auszulösen, seinen Körper zu reinigen, ein Impuls, der auch »Lady-Macbeth-Effekt« genannt wird.[13]

Die Reinigung bezieht sich ganz spezifisch auf die Körperteile, die an einem sündhaften Tun beteiligt waren. Teilnehmer an einem Experiment wurden dazu gebracht, eine imaginäre Person entweder am Telefon oder in einer E-Mail zu »belügen«. Bei einem anschließenden Test mit verschiedenen Produkten zogen Probanden, die am Telefon gelogen hatten, Mundwasser Seife vor, während diejenigen, die dies in einer E-Mail getan hatten, Seife Mundwasser vorzogen.[14]

Wenn ich Priming-Studien vor einem größeren Publikum vorstelle, ist die Reaktion oftmals Unglaube. Dies ist nicht weiter verwunderlich: System 2 glaubt, dass es die Kontrolle hat und dass es die Gründe für seine Entscheidungen kennt. Auch Ihnen drängen sich vermutlich Fragen auf: Wie ist es möglich, dass solche trivialen Manipulationen des Kontextes so weitreichende Auswirkungen haben? Beweisen diese Experimente, dass wir auf Gedeih und Verderb den primenden Reizen ausgeliefert sind, die sich zu jedem beliebigen Zeitpunkt in unserem Umfeld befinden? Natürlich nicht. Die Auswirkungen der Primes sind durchgängig nachweisbar, aber nicht unbedingt groß. Unter hundert Wählern werden nur einige wenige, deren anfängliche Präferenzen unsicher waren, in einer Schulfrage anders abstimmen, wenn sich ihr Wahllokal in einer Schule statt in einer Kirche befindet – aber ein paar Prozent können eine Wahl entscheiden.

Allerdings sind Zweifel hier nicht angebracht. Die Ergebnisse sind nicht erfunden, und sie sind auch keine statistischen Zufallsergebnisse. Sie müssen sich damit abfinden, dass die zentralen Schlussfolgerungen dieser Studien wahr sind. Wichtiger noch ist, dass Sie sich damit abfinden müssen, dass sie wahr in Bezug auf *Sie* sind. Wenn man Ihnen einen Bildschirmschoner mit schwimmenden Dollarscheinen dargeboten hätte,

hätten auch Sie wahrscheinlich weniger Bleistifte aufgehoben, um einem unbeholfenen Fremden zu helfen. Sie glauben nicht, dass diese Ergebnisse für Sie gelten, weil sie nicht Ihrem subjektiven Erleben entsprechen. Aber Ihr subjektives Erleben besteht größtenteils aus der Geschichte, die sich Ihr System 2 selbst über das erzählt, was vor sich geht. Priming-Phänomene entstehen in System 1, und Sie haben keinen bewussten Zugang dazu.

Ich beende dieses Kapitel mit einer perfekten Demonstration eines Priming-Effekts, die in der Büroküche einer britischen Universität stattfand.[15] Viele Jahre lang hatten Mitglieder dieses Büros den Tee oder Kaffee bezahlt, den sie sich während des Tages genommen hatten, indem sie Münzen in eine »Vertrauenskasse« warfen. Eine Liste mit Preisvorschlägen war ausgehängt. Eines Tages wurde ohne Vorwarnung oder Erklärung ein Poster über der Preisliste zur Schau gestellt. Zehn Wochen lang wurde jede Woche ein neues Bild gezeigt – entweder Blumen oder Augen, die den Betrachter direkt anzublicken schienen. Niemand sagte etwas zu dem neuen Wandschmuck, aber die Beiträge zur Vertrauenskasse veränderten sich deutlich. Die Poster und die Geldbeträge, die in die Sparbüchse geworfen wurden (im Verhältnis zu der verbrauchten Menge), sind in Abbildung 4 gezeigt. Sie verdienen eine genauere Betrachtung.

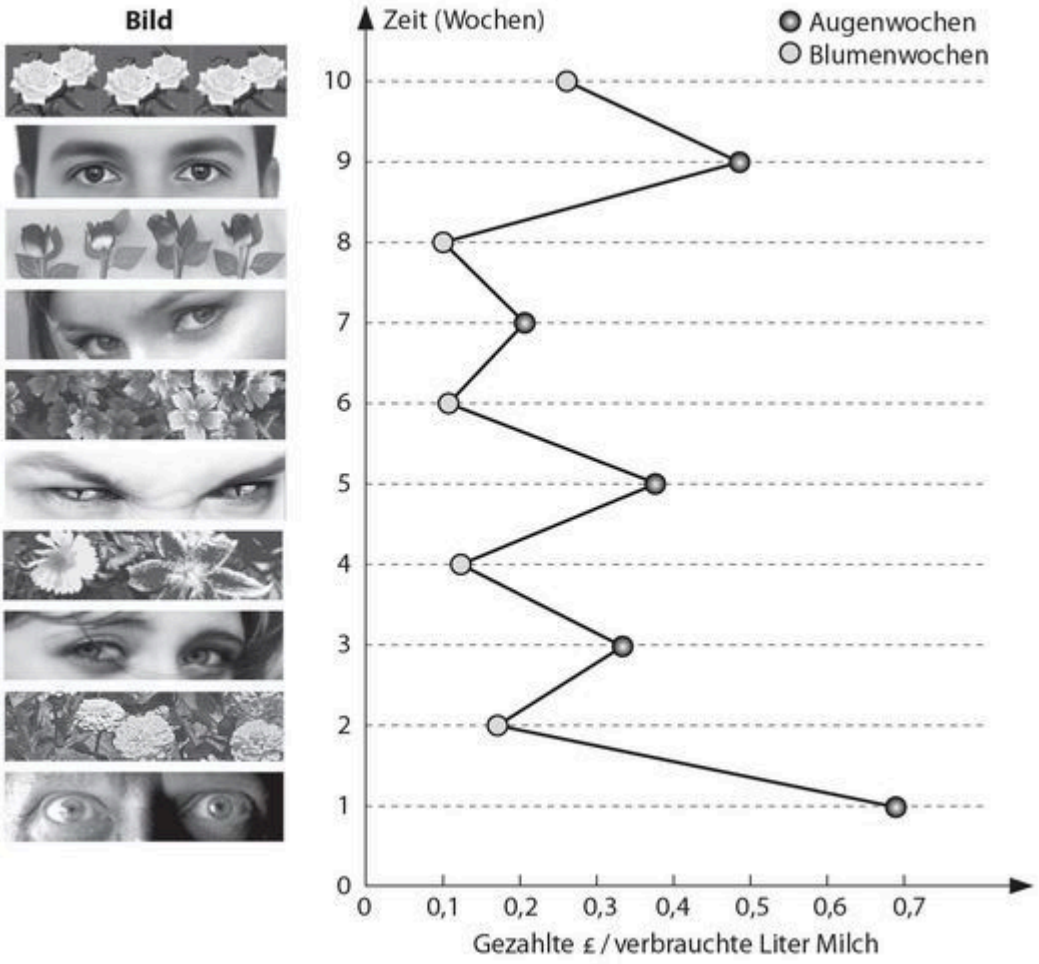


**Abbildung 4**

In der ersten Woche des Experiments (ganz unten in der Abbildung) starren zwei weit aufgerissene Augen die Kaffee- oder Teetrinker an, die durchschnittlich 70 Pence pro Liter Milch bezahlten. In der zweiten Woche zeigt das Poster Blumen, und die durchschnittlichen Beiträge fallen auf etwa 15 Pence. Der Trend hält an. Im Durchschnitt zahlten die Benutzer der Kaffeeküche in den »Augenwochen« fast dreimal so viel wie in den »Blumenwochen«. Eine rein symbolische Mahnung daran, unter Beobachtung zu stehen, veranlasste die Konsumenten offensichtlich dazu, sich »besser« zu benehmen. Wie wir es mittlerweile erwarten würden, tritt dieser Effekt ein, ohne dass es den Betreffenden auch nur ansatzweise bewusst wäre. Glauben Sie jetzt, dass Sie in das gleiche Muster verfallen würden?

Vor einigen Jahren schrieb der Psychologe Timothy Wilson ein Buch mit dem aufrüttelnden Originaltitel *Strangers to Ourselves* (»Uns selbst

fremd«).[16] Sie wurden jetzt mit diesem Fremden in Ihnen bekannt gemacht, der einen Großteil dessen kontrollieren mag, was Sie tun, auch wenn Sie dies nur selten erahnen. System 1 liefert die Eindrücke, die oftmals zu unseren Überzeugungen werden, und es ist die Quelle der Impulse, die häufig unsere Entscheidungen und Handlungen bestimmen. Es stellt eine implizite Interpretation dessen bereit, was uns widerfährt und was um uns herum geschieht, wobei es die Gegenwart mit der jüngsten Vergangenheit und mit Erwartungen über die nahe Zukunft verknüpft. Es enthält das Weltmodell, das Ereignisse sofort als normal oder überraschend bewertet. Es ist die Quelle unserer raschen und oftmals präzisen intuitiven Urteile. Und das meiste davon tut es, ohne dass wir uns seiner Aktivitäten bewusst wären. Wie wir in den folgenden Kapiteln sehen werden, ist System 1 auch der Ursprung vieler der systematischen Fehler in unseren Intuitionen.

### Zum Thema »Priming«

»Der Anblick all dieser Menschen in Uniform primt nicht Kreativität.«

»Die Welt ist bei Weitem nicht so sinnhaft aufgebaut, wie wir glauben. Die Kohärenz ist vor allem ein Produkt der Funktionsweise unseres Intellekts.«

»Sie wurden geprimt, Fehler zu finden, und genau das fanden sie.«

»Sein System 1 konstruierte eine Geschichte, und sein System 2 glaubte sie. Das passiert uns allen.«

»Ich brachte mich selbst zum Lächeln, und jetzt fühle ich mich tatsächlich besser!«

# *5. Kognitive Leichtigkeit*

Wenn Sie bei Bewusstsein sind – und vielleicht auch, wenn Sie es nicht sind –, laufen viele Berechnungen in Ihrem Gehirn ab, die gegenwärtige Antworten auf einige Schlüsselfragen aufrechterhalten und aktualisieren: Geschieht etwas Neues? Besteht eine Bedrohung? Läuft alles gut? Sollte meine Aufmerksamkeit neu ausgerichtet werden? Muss ich mich für diese Aufgabe stärker anstrengen? Stellen Sie sich ein Cockpit mit einer Reihe von Instrumenten vor, die die aktuellen Werte jeder dieser wichtigen Variablen anzeigen. Die Bewertungen werden automatisch von System 1 durchgeführt, und eine ihrer Aufgaben besteht darin, zu bestimmen, ob zusätzliche Anstrengungen von System 2 erforderlich sind.

Eine der Anzeigen misst die »kognitive Leichtigkeit« *(cognitive ease)*, und ihre Spannweite reicht von »hoch« bis »gering«.[1] Eine hohe Leichtigkeit ist ein Zeichen dafür, dass alles gut läuft – keine Bedrohungen, keine wichtigen Neuigkeiten, keine Notwendigkeit, die Aufmerksamkeit neu auszurichten oder sich stärker anzustrengen. Geringe Leichtigkeit deutet darauf hin, dass ein Problem existiert, welches eine verstärkte Mobilisierung von System 2 erfordert. Entsprechend erlebt man eine starke kognitive Leichtigkeit. Die »kognitive Beanspruchung« *(cognitive strain)* wird sowohl vom laufenden kognitiven Arbeitsaufwand als auch vom Vorhandensein unbefriedigter Bedürfnisse beeinflusst. Das Erstaunliche ist, dass eine einzige Anzeige der kognitiven Leichtigkeit mit einem großen Netzwerk unterschiedlichster Inputs und Outputs verknüpft ist.[2] Abbildung 5 fasst dies schematisch zusammen.

Das Schaubild legt die Vermutung nahe, dass ein Satz, der in einer deutlichen Schriftart gedruckt ist oder wiederholt wurde oder auch geprimt wurde, mit kognitiver Leichtigkeit flüssig verarbeitet wird. Auch wenn man, gut gelaunt, einem Sprecher zuhört oder wenn man einen Bleistift quer in den Mund gesteckt hat, sodass man »lachen« muss, erzeugt dies kognitive Leichtigkeit.

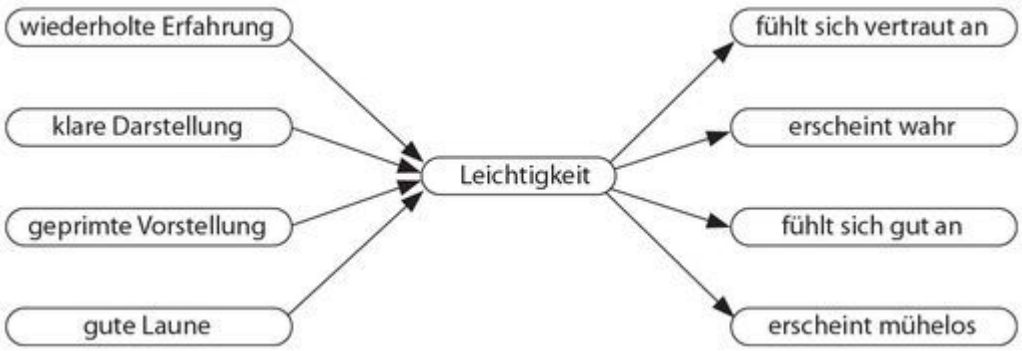


**Abbildung 5**
**Ursachen und Folgen kognitiver Leichtigkeit**

Die verschiedenen Ursachen für Leichtigkeit oder Anspannung haben austauschbare Wirkungen. Wenn Sie sich in einem Zustand kognitiver Leichtigkeit befinden, sind Sie vermutlich gut gelaunt, Sie mögen das, was Sie sehen, glauben das, was Sie hören, vertrauen Ihren Intuitionen und haben das Gefühl, dass Ihnen die gegenwärtige Situation in einer angenehmen Weise vertraut ist. Auch wird Ihr Denken vermutlich relativ beiläufig und oberflächlich sein. Wenn Sie angespannt sind, sind Sie vermutlich eher wachsam und argwöhnisch, strengen sich stärker an bei dem, was Sie tun, fühlen sich unbehaglicher und machen weniger Fehler, aber Sie sind auch weniger intuitiv und weniger kreativ als sonst.

## Illusionen des Gedächtnisses

Bei dem Wort »Illusion« denkt man unwillkürlich an optische Täuschungen, weil wir alle Bilder kennen, die uns in die Irre führen. Aber das Sehen ist nicht der einzige Bereich, in dem Illusionen auftreten; auch das Gedächtnis unterliegt Täuschungen, so wie das Denken im Allgemeinen.

David Stenbill, Monica Bigoutski, Shana Tirana. Ich habe diese Namen gerade erfunden. Wenn Sie irgendeinem davon in den nächsten Minuten begegnen, werden Sie sich wahrscheinlich daran erinnern, wo Sie die Namen zum ersten Mal gesehen haben. Sie wissen – und Sie werden schon seit einiger Zeit wissen –, dass dies nicht die Namen von B-Promis sind. Aber nehmen wir an, dass Ihnen in ein paar Tagen eine lange Liste mit Namen gezeigt wird, einschließlich einiger Namen von B-Promis und mit »neuen« Namen von Menschen, von denen Sie noch nie gehört haben; Ihre Aufgabe ist es nun, jeden Prominentennamen auf der Liste zu überprüfen. Es besteht eine erhebliche Wahrscheinlichkeit, dass Sie David Stenbill als einen Prominenten identifizieren werden, auch wenn Sie (selbstverständlich) nicht wissen werden, ob Ihnen dieser Name im Zusammenhang mit Kinofilmen, Sport oder Politik begegnet ist. Larry Jacoby, der Psychologe, der diese Gedächtnistäuschung als Erster im Labor nachgewiesen hat, gab seinem Artikel den Titel »Über Nacht berühmt werden«.[3] Wie ist dies zu erklären? Beginnen Sie damit, sich zu fragen, woher Sie wissen, ob jemand berühmt ist oder nicht. In einigen Fällen wirklich berühmter Menschen (beziehungsweise von Prominenten auf einem Feld, das Sie interessiert) besitzen Sie vielleicht eine mentale Akte mit reichhaltigen Informationen über eine Person – denken Sie nur an Albert Einstein, Bono oder Hillary Clinton. Aber Sie besitzen keine Akte mit Informationen über David Stenbill, wenn Sie seinem Namen in ein paar Tagen begegnen. Sie haben lediglich ein Gefühl der Vertrautheit – irgendwo haben Sie diesen Namen schon einmal gesehen.

Jacoby formulierte das Problem in hübschen Worten: »Die Erfahrung der Vertrautheit zeichnet sich durch eine einfache, aber starke Qualität der ›Vergangenheit‹ aus, die anzuzeigen scheint, dass sie die direkte

Widerspiegelung einer früheren Erfahrung ist.«[4] Diese Qualität der Vergangenheit ist eine Illusion. Wie Jacoby und viele Wissenschaftler nach ihm gezeigt haben, ist es tatsächlich so, dass der Name David Stenbill Ihnen vertraut vorkommen wird, wenn Sie ihn sehen, *weil Sie ihn deutlicher sehen*. Wörter, die man schon einmal gesehen hat, sind leichter wiederzuerkennen – man kann sie besser identifizieren als andere Wörter, wenn sie sehr kurz dargeboten oder durch Rauschen überdeckt werden, und man kann sie (um ein paar Hundertstelsekunden) schneller lesen als andere Wörter. Kurzum, man erlebt eine größere kognitive Leichtigkeit, wenn man ein Wort wahrnimmt, das man schon einmal gesehen hat, und es ist dieses Gefühl der Mühelosigkeit, das den Eindruck der Vertrautheit hervorruft.

Abbildung 5 legt eine Methode nahe, um dies zu überprüfen. Wählen Sie ein völlig neues Wort aus, machen Sie es klar und deutlich sichtbar, und es wird eher die Qualität der Vergangenheit besitzen. Tatsächlich wird ein neues Wort eher als vertraut erkannt, wenn es unbewusst geprimt wird, indem man es einige Millisekunden vor dem Test darbietet oder es kontrastreicher zeigt als einige andere Wörter in der Liste. Der Zusammenhang funktioniert auch in die andere Richtung. Stellen Sie sich vor, man zeigt Ihnen eine Liste mit Wörtern, die mehr oder minder unscharf sind. Einige der Wörter sind total verschwommen, andere sind es in geringerem Maße, und Ihre Aufgabe besteht darin, die Wörter zu erkennen, die deutlicher dargeboten werden. Ein Wort, das Sie vor Kurzem gesehen haben, wird Ihnen deutlicher vorkommen als unbekannte Wörter. Der Abbildung 5 lässt sich entnehmen, dass die verschiedenen Wege, die kognitive Beanspruchung zu verringern oder zu erhöhen, gegeneinander austauschbar sind; Sie mögen dabei nicht genau wissen, wieso die kognitive Beanspruchung bald geringer, bald stärker ist. Auf diese Weise entsteht die Illusion der Vertrautheit.

## Illusionen der Wahrheit

»New York ist eine Großstadt in den Vereinigten Staaten.« »Der Mond dreht sich um die Erde.« »Ein Huhn hat vier Beine.« In all diesen Fällen haben Sie schnell eine Menge damit zusammenhängender Informationen abgerufen, die fast alle in die eine oder andere Richtung weisen. Sie wussten gleich nach der Lektüre, dass die ersten beiden Aussagen wahr sind, während die letzte falsch ist. Nun ist es allerdings so, dass die Aussage »Ein Huhn hat drei Beine« offensichtlicher falsch ist als »Ein Huhn hat vier Beine«. Ihre Assoziationsmaschinerie verlangsamt das Urteil über den letzten Satz, indem sie die Tatsache in Erinnerung ruft, dass viele Tiere vier Beine haben, und vielleicht auch, dass Supermärkte oftmals Hähnchenschlegel in Viererpackungen verkaufen. System 2 war an der Überprüfung dieser Informationen beteiligt und hatte vielleicht Bedenken, ob die Frage über New York zu leicht war, oder es überprüfte die Bedeutung von »sich drehen«.

Denken Sie an Ihre Führerscheinprüfung. Stimmt es, dass Sie für das Fahren eines Fahrzeugs, das über drei Tonnen wiegt, eine Sondererlaubnis brauchen? Vielleicht haben Sie sich ernsthaft auf die Prüfung vorbereitet, und Sie erinnern sich daran, ob die Antwort links oder rechts auf der Seite auftauchte und welche Erklärung dafür gegeben wurde. Allerdings bin ich ganz anders an meine Fahrprüfungen herangegangen, wenn ich in einen neuen US-Bundesstaat umgezogen bin. Ich habe die Broschüre mit den Verkehrsregeln einmal kurz durchgelesen und dann das Beste gehofft. Aufgrund meiner langjährigen Fahrpraxis kannte ich einige der Antworten. Aber es gab Fragen, auf die mir keine gute Antwort einfiel und wo ich als Richtschnur nur die kognitive Leichtigkeit hatte. Wenn sich die Antwort vertraut anhörte, nahm ich an, dass sie vermutlich wahr war. Wenn sie mir neu vorkam (oder äußerst ungewöhnlich), verwarf ich sie. Der Eindruck der Vertrautheit wird von System 1 erzeugt, und System 2 stützt sich auf diesen Eindruck für ein Wahr-falsch-Urteil.[5]

Die Lektion von Abbildung 5 lautet, dass Illusionen unvermeidlich sind, wenn ein Urteil auf den Eindruck geringer oder starker kognitiver Beanspruchung gestützt wird. Alles, was das reibungslose Funktionieren der Assoziationsmaschine erleichtert, wird auch Überzeugungen verzerren.

Eine zuverlässige Methode, Menschen dazu zu bringen, falsche Aussagen zu glauben, ist häufiges Wiederholen, weil Vertrautheit sich nicht leicht von Wahrheit unterscheiden lässt. Auch autoritäre Institutionen und Marketing-Spezialisten wissen das seit jeher. Aber Psychologen haben herausgefunden, dass man die Tatsache oder Idee nicht vollständig wiederholen muss, um ihr den Anschein der Wahrheit zu geben. Personen, denen wiederholt der Ausdruck »die Körpertemperatur eines Huhns« dargeboten wurde, akzeptierten die Aussage »die Körpertemperatur eines Huhns beträgt sechzig Grad Celsius« (oder irgendeine andere beliebige Zahl) eher als wahr.[6] Die Vertrautheit mit einem Ausdruck innerhalb der Aussage reichte aus, um die ganze Aussage vertraut und damit wahr erscheinen zu lassen. Wenn man sich nicht an die Quelle einer Aussage erinnern kann und keine Möglichkeit hat, sich auf andere Dinge zu beziehen, die man kennt, bleibt einem nichts anderes übrig, als sich an die gefühlte Mühelosigkeit des Denkens zu halten.

## Wie man eine überzeugende Mitteilung schreibt

Angenommen, Sie müssen eine Mitteilung schreiben, die die Empfänger unbedingt glauben sollen. Selbstverständlich wird Ihre Mitteilung wahr sein, aber das genügt nicht unbedingt, um Menschen dazu zu veranlassen, sie für wahr zu halten. Es ist völlig legitim, wenn Sie sich den Trick der kognitiven Leichtigkeit gezielt zunutze machen, und Studien über »Wahrheitsillusionen« liefern konkrete Anhaltspunkte, die Ihnen dabei helfen können, dieses Ziel zu erreichen.

Der allgemeine Grundsatz lautet, dass alles, was Sie tun können, um die kognitive Beanspruchung des Lesers zu verringern, hilfreich ist. Also sollten Sie als Erstes Ihre Mitteilung so lesbar wie möglich machen. Vergleichen Sie diese beiden Aussagen:

**Adolf Hitler wurde 1892 geboren.**
Adolf Hitler wurde 1887 geboren.

Beide sind falsch (Hitler wurde 1889 geboren), aber Experimente haben gezeigt, dass die erste Aussage eher geglaubt wird. Ein weiterer Rat: Wenn Ihre Mitteilung gedruckt werden soll, verwenden Sie hochwertiges Papier, um den Kontrast zwischen den Buchstaben und dem Hintergrund zu verstärken. Falls Sie Farbe verwenden, wird man Ihnen eher glauben, wenn Ihr Text leuchtend blau oder rot gedruckt ist, als in mittleren Grün- oder Gelbtönen oder in Blassblau.

Wenn Sie Wert darauf legen, für glaubwürdig und intelligent gehalten zu werden, sollten Sie sich nicht kompliziert ausdrücken, wenn man das Gleiche auch in einfachen Worten sagen kann. Mein Kollege an der Universität Princeton, Danny Oppenheimer, widerlegte einen bei Studenten weitverbreiteten Mythos darüber, welches Vokabular Professoren am eindrucksvollsten finden. In einem Artikel mit dem Titel »Consequences of Erudite Vernacular Utilized Irrespective of Necessity: Problems with Using Long Words Needlessly« (»Folgen des Gebrauchs von Fachjargon ungeachtet der Notwendigkeit: Probleme, die dadurch entstehen, dass man

unnötigerweise lange Wörter benutzt«) zeigte er, dass es ein Anzeichen geringer Intelligenz und Glaubwürdigkeit ist, wenn man geläufige Ideen in einer hochtrabenden Sprache formuliert.[7]

Sie sollten Ihre Mitteilung aber nicht nur möglichst einfach, sondern auch einprägsam formulieren. Drücken Sie Ihre Gedanken möglichst in gebundener Sprache aus; sie werden dann eher für wahr gehalten. Die Probanden in einem viel beachteten Experiment lasen Dutzende wenig bekannter Sentenzen wie:

> *Woes unite foes.* (Die Not verbindet Feinde.)
> *Little strokes will tumble great oaks.* (Steter Tropfen höhlt den Stein.)
> *A fault confessed is half redressed.* (Einsicht ist der erste Schritt zur Besserung.)

Andere Studenten lasen einige gleichbedeutende Sentenzen, die in nicht reimende Versionen umformuliert worden waren:

> *Woes unite enemies.*
> *Little strokes will tumble great trees.*
> *A fault admitted is half redressed.*

Die gereimten Aphorismen wurden als sinnreicher beurteilt als ihre nicht gereimten Pendants.[8]

Schließlich sollten Sie dann, wenn Sie eine Quelle zitieren, eine mit einem Namen wählen, der sich leicht aussprechen lässt. Die Teilnehmer an einem Experiment sollten die Geschäftsaussichten fiktiver türkischer Unternehmen auf der Basis der Analysen zweier Brokerfirmen beurteilen.[9] Für jede Aktie stammte einer der Berichte von einer Firma mit einem leicht auszusprechenden Namen (z. B. Artan), während der andere Bericht von einer Firma mit schwer auszusprechendem Namen (z. B. Taahhut) stammte.

Die Berichte widersprachen einander manchmal. Die beste Vorgehensweise für die Beobachter hätte darin bestanden, beide Berichte zu mitteln, aber das taten sie nicht. Sie gaben dem Bericht von Artan viel mehr Gewicht als dem von Taahhut. Erinnern Sie sich daran, dass System 2 faul ist und dass geistige Anstrengung gemieden wird. Wenn möglich, wollen die Empfänger Ihrer Nachricht sich von allem fernhalten, was sie an Anstrengung erinnert, und dazu gehört auch eine Quelle mit einem komplizierten Namen.

All dies sind sehr gute Ratschläge, aber wir sollten es nicht übertreiben. Hochwertiges Papier, strahlende Farben und Reime oder einfache Ausdrucksweise werden nicht besonders hilfreich sein, wenn Ihre Mitteilung offenkundig unsinnig ist oder wenn sie Tatsachen widerspricht, deren Wahrheit der Empfänger der Nachricht kennt. Die Psychologen, die diese Experimente durchführen, halten Menschen nicht für dumm oder unendlich leichtgläubig. Psychologen sind vielmehr der Ansicht, dass wir alle uns während eines Großteils unseres Lebens von den Eindrücken von System 1 leiten lassen – und wir kennen oft nicht die Quelle dieser Eindrücke. Woher wissen wir, dass eine Aussage wahr ist? Wenn sie logisch oder assoziativ eng mit anderen Überzeugungen oder Präferenzen von uns verbunden ist oder wenn sie von einer Quelle stammt, der wir vertrauen und die wir mögen, erleben wir eine Art kognitive Leichtigkeit. Das Problem ist, dass dieses Gefühl der Mühelosigkeit auch andere Ursachen haben kann – etwa die Qualität der Schriftart oder der eingängige sprachliche Rhythmus – und dass es keine einfache Methode gibt, um Gefühle auf ihre wahren Ursachen zurückzuführen. Dies ist die Botschaft von Abbildung 5: Das Gefühl der Leichtigkeit oder Beanspruchung hat vielfältige Ursachen, und diese lassen sich nur schwer auftrennen. Es ist schwer, aber nicht unmöglich. Menschen können einige der oberflächlichen Faktoren, die Wahrheitsillusionen erzeugen, überwinden, wenn sie stark dazu motiviert sind. Doch bei vielen Gelegenheiten wird das faule System 2 den Vorschlägen von System 1 folgen und weiterziehen.

## Beanspruchung und Anstrengung

Die Symmetrie vieler assoziativer Verknüpfungen war ein beherrschendes Thema in der Diskussion der assoziativen Kohärenz. Wie wir zuvor gesehen haben, neigen Menschen, die dadurch, dass sie einen Bleistift in ihren Mund stecken oder eine Kugel zwischen ihren zusammengepressten Brauen halten, ihr Gesicht unwillkürlich zu einem Lächeln oder Stirnrunzeln verziehen, dazu, die Emotionen zu erleben, die Stirnrunzeln und Lächeln normalerweise ausdrücken. Der gleichen selbstverstärkenden Wechselseitigkeit begegnet man in Studien über kognitive Leichtigkeit. Einerseits erleben wir eine starke Beanspruchung unserer kognitiven Leistungsfähigkeit, wenn die anstrengenden Operationen von System 2 ablaufen. Andererseits mobilisiert eine hohe kognitive Beanspruchung unabhängig von ihrer Ursache tendenziell System 2, mit der Folge, dass von einem beiläufigen, intuitiven Modus der Problemlösung auf einen konzentrierten, analytischen Modus umgestellt wird.[10]

Das Schläger-und-Ball-Problem wurde bereits als ein Test für die Neigung von Menschen beschrieben, Fragen mit der ersten Idee zu beantworten, die ihnen einfällt, ohne diese zu überprüfen. Der von Shane Frederick entwickelte *Cognitive Reflection Test* (CRT) besteht aus dem Schläger-und-Ball-Problem und zwei weiteren, die alle ausgewählt wurden, weil sie eine sofortige intuitive Antwort hervorrufen, die falsch ist. Die zwei anderen Denkaufgaben im CRT sind:

> Wenn 5 Maschinen 5 Minuten brauchen, um 5 Geräte herzustellen, wie lange brauchen dann 100 Maschinen, um 100 Geräte herzustellen?
>
> 100 Minuten oder 5 Minuten?

> In einem See breitet sich ein kleines Feld von Seerosen aus. Jeden Tag verdoppelt sich die Größe des Feldes. Wenn es 48

Tage dauert, bis die Seerosen den ganzen See bedecken, wie lange dauert es dann, bis sie die Hälfte des Sees bedecken?

24 Tage oder 47 Tage?

Die richtigen Antworten finden Sie in der Anmerkung.[11] Vierzig Studenten der Universität Princeton wurden dem Test unterzogen. Die Hälfte von ihnen sah die Aufgaben in einer kleinen Schrift und in einem ausgewaschenen grauen Druckbild. Die Aufgaben waren lesbar, aber die Schrift erzeugte eine hohe kognitive Beanspruchung. Die Ergebnisse lassen an Klarheit nichts zu wünschen übrig: 90 Prozent der Studenten, die den CRT in normaler Schrift sahen, machten bei dem Test wenigstens einen Fehler; während dieser Prozentsatz auf 35 Prozent fiel, wenn die Schrift kaum lesbar war. Sie haben richtig gelesen: Bei der schlechten Schrift war die Leistung besser. Eine starke kognitive Beanspruchung mobilisiert, unabhängig von ihrer Ursache, System 2, das die von System 1 vorgeschlagene intuitive Antwort eher verwirft.

## Die Freuden mühelosen Denkens

Ein Artikel mit dem Titel »Mentale Entspannung lässt uns lächeln« beschreibt ein Experiment, bei dem den Teilnehmern kurz Bilder von Gegenständen gezeigt wurden.[12] Bei einigen dieser Bilder wurde die Erkennbarkeit des Objekts dadurch erleichtert, dass seine Silhouette dargeboten wurde, unmittelbar bevor das vollständige Bild zu sehen war – allerdings so kurz, dass die Konturen nicht bewusst wahrgenommen werden konnten. Emotionale Reaktionen wurden durch Aufzeichnung elektrischer Impulse an den Gesichtsmuskeln gemessen, womit sich Veränderungen der Mimik registrieren lassen, die so schwach und so kurz sind, dass sie Beobachtern entgehen. Erwartungsgemäß zeigten die Probanden ein leises Lächeln und entspannte Brauen, wenn die Bilder leichter zu erkennen waren. Es scheint eine Eigenschaft von System 1 zu sein, dass kognitive Leichtigkeit mit guten Gefühlen einhergeht.

Wie erwartet, rufen leicht auszusprechende Wörter eine wohlwollende Einstellung hervor. Unternehmen mit unkomplizierten Namen werden in der ersten Woche nach ihrem Börsengang am Markt höher bewertet als andere, auch wenn dieser Effekt mit der Zeit verschwindet. Aktien mit leicht auszusprechenden Handelssymbolen (wie KAR oder LUNMOO) haben eine bessere Wertentwicklung als diejenigen mit zungenbrecherischen Kurzbezeichnungen wie PXG oder RDO – und sie scheinen diesen kleinen Vorteil auch längerfristig zu behalten. [13] Bei einer Studie in der Schweiz kam heraus, dass Investoren glauben, Aktien mit flüssigen Namen wie Emmi, Swissfirst und Comet würfen höhere Renditen ab als solche mit sperrigen Bezeichnungen wie Geberit und Ypsomed.[14]

Wie wir in Abbildung 5 sahen, führt Wiederholung zu kognitiver Leichtigkeit und einem angenehmen Gefühl der Vertrautheit. Der berühmte Psychologe Robert Zajonc widmete einen Großteil seiner wissenschaftlichen Laufbahn der Erforschung des Zusammenhangs zwischen der Wiederholung eines beliebigen Reizes und der leichten positiven Einstellung, die Menschen, die diesem Reiz ausgesetzt wurden, dazu entwickeln. Zajonc nannte dieses Phänomen »Mere-Exposure-Effekt«

(»Effekt der bloßen Darbietung«).[15] Ein Experiment, das mithilfe der Studentenzeitungen der Universität von Michigan und der Staatsuniversität von Michigan durchgeführt wurde, ist eines meiner Lieblingsexperimente.[16] Über mehrere Wochen hinweg erschien auf der Titelseite der Zeitungen ein Kästchen, ähnlich wie bei einer Werbeanzeige, das die folgenden türkischen (oder zumindest so klingenden) Wörter enthielt: *kadirga*, *saricik*, *biwonjni*, *nansoma* und *iktitaf*. Die Häufigkeit, mit der die Wörter wiederholt wurden, schwankte: Eines der Wörter wurde nur einmal gezeigt, die anderen tauchten bei 2, 5, 10 oder 25 verschiedenen Gelegenheiten auf. (Die Wörter, die in einer der Studentenzeitungen am häufigsten dargeboten wurden, waren in der anderen Zeitung die am seltensten gezeigten.) Es wurde keine Erklärung angegeben, und Fragen von Lesern wurden mit der Aussage beantwortet: »Der Käufer der Anzeige wünscht Anonymität.«

Als die rätselhafte Anzeigenserie endete, schickten die Studienleiter Fragebogen an die Studenten der Universitäten, in denen sie diese nach ihren Eindrücken fragten, ob jedes der Wörter »etwas Gutes oder etwas Schlechtes bedeutet«. Die Ergebnisse waren spektakulär: Die häufiger dargebotenen Wörter wurden viel positiver bewertet als die Wörter, die nur ein- oder zweimal gezeigt wurden. Das Ergebnis wurde in zahlreichen Experimenten bestätigt, in denen unter anderem chinesische Ideogramme, Gesichter und zufällig geformte Polygone verwendet wurden.

Der Mere-Exposure-Effekt hängt nicht vom bewussten Erleben des Bekanntseins ab. Tatsächlich ist der Effekt nicht einmal auf bewusste Wahrnehmung angewiesen: Er tritt selbst dann auf, wenn die wiederholten Wörter oder Bilder so schnell gezeigt werden, dass es den Beobachtern nicht bewusst wird, sie überhaupt gesehen zu haben. Trotzdem haben sie eine positivere Einstellung zu den Wörtern oder Bildern, die häufiger dargeboten wurden. Wie mittlerweile klar sein sollte, kann System 1 auf Eindrücke von Ereignissen reagieren, die System 2 nicht bewusst registriert hat. Tatsächlich ist der Effekt der bloßen Darbietung bei Stimuli, die nicht bewusst wahrgenommen werden, sogar noch stärker.[17]

Zajonc behauptete, der Effekt der Wiederholung auf die Einstellung sei eine zutiefst biologisch verankerte Tatsache, und er gelte für alle Tiere. Um in einer oftmals gefährlichen Welt zu überleben, sollte ein Organismus vorsichtig auf einen neuartigen Reiz reagieren, mit Rückzug und Furcht. Die Überlebenschancen eines Tieres, das vor neuartigen Reizen nicht auf

der Hut ist, sind gering. Aber es ist auch adaptiv, wenn die anfängliche Vorsicht nachlässt, sobald sich zeigt, dass der Reiz keine Bedrohung darstellt. Der Mere-Exposure-Effekt lässt sich laut Zajonc damit erklären, dass auf die wiederholte Darbietung eines Reizes keine negative Erfahrung folgt. Ein solcher Reiz wird langfristig zu einem Sicherheitssignal, und Sicherheit ist positiv. Dieses Anpassungsmuster ist offenkundig nicht auf Menschen beschränkt. Um dies zu beweisen, setzte einer von Zajoncs Mitarbeitern zwei Gruppen befruchteter Hühnereier verschiedenen Tönen aus. Nachdem die Küken geschlüpft waren, stießen sie beständig weniger negative Stresssignale aus, wenn sie dem Ton ausgesetzt wurden, den sie gehört hatten, als sie noch im Ei waren.[18]

Zajonc gab eine beredte Zusammenfassung seines Forschungsprojekts:

> Die Folgen wiederholter Expositionen nützen dem Organismus in seinen Beziehungen zur unmittelbaren belebten und unbelebten Umwelt. Sie erlauben es dem Organismus, zwischen sicheren und unsicheren Objekten und Habitaten zu unterscheiden, und sie sind die elementarste Grundlage sozialer Bindungen. Daher bilden sie auch die Basis der sozialen Organisation und des sozialen Zusammenhalts – und diese sind die elementaren Quellen psychischer und sozialer Stabilität.[19]

Der Zusammenhang zwischen positiver Gestimmtheit und kognitiver Leichtigkeit in System 1 hat eine lange evolutionäre Geschichte.

## Leichtigkeit, Stimmung und Intuition

Um 1960 herum glaubte ein junger Psychologe namens Sarnoff Mednick, er habe das Wesen der Kreativität ergründet. Seine Idee war ebenso einfach wie überzeugend: Kreativität basiert auf einem außergewöhnlich gut funktionierenden assoziativen Gedächtnis. Er entwickelte einen Test, den sogenannten *Remote Association Test* (RAT), der noch immer in Kreativitätsstudien vielfache Anwendung findet.

Ein einfaches Beispiel liefern die folgenden drei Wörter:

| Hütte | Schweiz | Kuchen |
|---|---|---|
| *(cottage)* | *(Swiss)* | *(cake)* |

Fällt Ihnen ein Wort ein, das mit all diesen drei Wörtern assoziiert ist? Vermutlich kommen Sie darauf, dass die Antwort »Käse« lautet. Versuchen Sie es jetzt bei folgender Kombination:

| Hechtsprung/Sturzflug | Licht | Rakete |
|---|---|---|
| *(dive)* | *(light)* | *(rocket)* |

Dieses Problem ist viel schwieriger, aber es hat nur eine richtige Antwort, die jeder Sprecher des Englischen erkennt, auch wenn weniger als 20 Prozent einer Stichprobe von Studenten sie innerhalb von 15 Sekunden fanden. Die Antwort lautet »Himmel«. Selbstverständlich hat nicht jede Worttriade eine Lösung. So haben beispielsweise die Wörter »Traum«,

»Ball«, »Buch« keine gemeinsame Assoziation, die von allen als gültig anerkannt wird.

Mehrere deutsche Forschergruppen, die in den letzten Jahren mit dem RAT arbeiteten, haben bemerkenswerte Erkenntnisse über die kognitive Leichtigkeit zutage gefördert. Eine der Gruppen wollte folgende Fragen klären: Spüren Menschen, dass eine Worttriade eine Lösung besitzt, bevor sie wissen, was diese Lösung ist? Wie beeinflusst die Stimmungslage die Leistung bei dieser Aufgabe? Um dies herauszufinden, veränderten sie die Stimmungslage ihrer Probanden zum Positiven oder zum Negativen, indem sie sie aufforderten, mehrere Minuten lang über glückliche oder traurige Episoden in ihrem Leben nachzudenken. Anschließend boten sie diesen Versuchspersonen eine Reihe von Dreiwortgruppen dar, die Hälfte davon assoziativ verknüpft (wie etwa »Hechtsprung«, »Licht«, »Rakete«) und die andere Hälfte nicht assoziativ verknüpft (wie etwa »Traum«, »Ball«, »Buch«), und wiesen sie an, sehr schnell eine von zwei Tasten zu drücken, um anzugeben, ob sie eine assoziative Verknüpfung vermuteten oder nicht.[20] Die erlaubte Zeitspanne für diese Einschätzung – zwei Sekunden – war viel zu kurz, um das Problem gründlich zu durchdenken und die richtige Antwort zu finden.

Die erste Überraschung bestand darin, dass die Vermutungen sehr viel öfter zutreffend waren, als dies bei einer reinen Zufallsauswahl der Fall gewesen wäre. Ich finde dies erstaunlich. Ein Gefühl kognitiver Leichtigkeit wird offenbar durch ein sehr schwaches Signal aus der Assoziationsmaschine erzeugt, die »weiß«, dass die drei Wörter kohärent sind (eine Assoziation gemeinsam haben) – lange bevor die Assoziation abgerufen wird.[21] Die Rolle kognitiver Leichtigkeit bei solchen Urteilsprozessen wurde von einer anderen deutschen Gruppe bestätigt: Manipulationen, die die Flüssigkeit der kognitiven Verarbeitung erhöhen (Priming, klare Schriftart, vorhergehende Darbietung von Wörtern), steigern die Neigung, assoziative Verknüpfungen zwischen Wörtern zu erkennen.[22]

Eine weitere bemerkenswerte Entdeckung ist der starke Einfluss der Stimmungslage auf diese intuitive Leistung. Die Experimentatoren berechneten einen »Intuitionsindex«, um die Treffgenauigkeit zu messen. Sie fanden heraus, dass sich die Treffgenauigkeit mehr als verdoppelte, wenn sie die Probanden vor dem Test in eine gute Stimmungslage

versetzten, indem sie ihnen positive Gedanken eingaben.[23] Noch bemerkenswerter war der Befund, dass deprimierte Probanden bei der intuitiven Aufgabe völlig versagten; ihre Vermutungen lagen innerhalb der statistischen Zufallserwartung. Die Stimmungslage wirkt sich offenkundig auf die Operation von System 1 aus: Wenn wir uns unwohl und unzufrieden fühlen, verlieren wir den Kontakt zu unserer Intuition.

Diese Ergebnisse fügen sich ein in die wachsende Zahl von empirischen Befunden, die darauf hindeuten, dass eine positive Stimmungslage, Intuition, Kreativität, Leichtgläubigkeit und zunehmende Beanspruchung von System 1 ein Cluster bilden.[24] Andererseits sind auch Niedergeschlagenheit, Vigilanz, Argwohn, eine analytische Herangehensweise und vermehrte Anstrengung eng miteinander verbunden. Eine fröhliche Stimmung lockert die Kontrolle von System 2 über die Leistung: Wenn wir gut gelaunt sind, werden wir intuitiver und kreativer, aber auch weniger aufmerksam und anfälliger für logische Fehler. Auch diese Verknüpfung ist, wie der Mere-Exposure-Effekt, biologisch sinnvoll. Gute Laune ist ein Signal, dass alles eher gut läuft, dass die Umgebung sicher ist und dass man nicht auf der Hut sein muss. Schlechte Laune deutet darauf hin, dass die Dinge nicht so gut laufen, dass es womöglich eine Bedrohung gibt und dass erhöhte Daueraufmerksamkeit erforderlich ist. Kognitive Leichtigkeit ist sowohl eine Ursache als auch eine Folge einer positiven Gemütslage.

Der *Remote Association Test* sagt uns noch mehr über den Zusammenhang zwischen geringer kognitiver Beanspruchung und einem positiven Affekt. Betrachten Sie kurz die beiden Worttriaden:

| | | |
|---|---|---|
| Schlaf | Post | Schalter |
| *(sleep)* | *(mail)* | *(switch)* |
| Salz | tief | Gischt |

*(salt)  (deep)  (foam)*

Sie bemerken nichts davon, aber Messungen der elektrischen Aktivität in Ihren Gesichtsmuskeln hätten vermutlich auf ein leichtes Lächeln hingedeutet, als Sie die zweite Dreiergruppe lasen, die kohärent ist (»Meer« ist hier die Lösung). Diese lächelnde Reaktion auf Kohärenz ist bei Probanden festzustellen, denen nichts über gemeinsame assoziative Termini gesagt wird; ihnen wird lediglich eine vertikal angeordnete Worttriade gezeigt, und man instruiert sie, die Leertaste zu drücken, nachdem sie diese gelesen haben.[25] Der Eindruck kognitiver Leichtigkeit, der mit der Präsentation einer kohärenten Triade verbunden ist, scheint an sich eine leicht lustvolle Erfahrung zu sein.

Die Daten, die wir über eine positive Stimmungslage, kognitive Leichtigkeit und die Intuition von Kohärenz besitzen, sind, wie die Wissenschaftler sagen, korrelativ – wechselseitig –, aber nicht unbedingt kausal. Mühelosigkeit der Kognition und Lächeln treten zusammen auf, aber ist die gute Stimmung Ursache von Intuitionen der Kohärenz? Ja. Der Beweis stammt aus einem raffinierten experimentellen Ansatz, der sich immer größerer Beliebtheit erfreut. Einigen Versuchspersonen wurde eine Deckgeschichte präsentiert, die eine alternative Interpretation ihrer positiven Stimmung nahelegte; man sagte ihnen über die Musik, die in ihren Kopfhörern lief: »Frühere Forschungen haben gezeigt, dass diese Musik die emotionalen Reaktionen von Individuen beeinflusst.« [26] Diese Geschichte löscht die Intuition der Kohärenz vollständig aus. Der Befund zeigt, dass die kurze emotionale Reaktion im Anschluss an die Darbietung einer Worttriade (eine angenehme Reaktion, wenn die Triade kohärent ist, andernfalls eine unangenehme) die eigentliche Basis von Kohärenzurteilen ist. Hier gibt es nichts, was System 1 nicht tun kann. Emotionale Veränderungen werden jetzt erwartet, und weil sie nicht überraschend sind, werden sie nicht kausal mit den Wörtern verbunden.

Dies sind Beispiele für hervorragende psychologische Studien, die verschiedene experimentelle Verfahren kombinieren und Befunde hervorbringen, die sowohl robust als auch höchst überraschend sind. In den letzten Jahrzehnten haben wir eine Menge über die automatischen

Funktionsmechanismen von System 1 gelernt. Ein Großteil dessen, was wir heute wissen, hätte sich vor dreißig oder vierzig Jahren noch wie Science-Fiction angehört. Es lag jenseits der Vorstellungskraft, dass eine undeutliche Schrift Wahrheitsurteile beeinflusst und die kognitive Leistungsfähigkeit steigert oder dass eine emotionale Reaktion auf die Leichtigkeit der kognitiven Verarbeitung einer Worttriade Kohärenzeindrücke vermittelt. Die Psychologie hat große Fortschritte gemacht.

## Zum Thema »Kognitive Leichtigkeit«

»Wir sollten den Geschäftsplan nicht bloß deshalb verwerfen, weil die Schriftart unleserlich ist.«

»Wir neigen dazu, es zu glauben, weil es so oft wiederholt wurde, aber wir sollten es noch einmal durchdenken.«

»Bekanntheit erzeugt eine positive Einstellung. Dies ist der Mere-Exposure-Effekt.«

»Ich bin heute in bester Laune, und mein System 2 ist schwächer als gewöhnlich. Ich sollte besonders sorgfältig sein.«

# *6. Normen, Überraschungen und Ursachen*

Wir haben jetzt die zentralen Merkmale und Funktionen von System 1 und System 2 kennengelernt, wobei System 1 detaillierter dargestellt wurde. Vereinfacht können wir sagen, dass in unserem Kopf ein bemerkenswert leistungsfähiger Computer sitzt, der nach herkömmlichen Hardware-Standards nicht schnell arbeitet, aber in der Lage ist, die Struktur unserer Welt mit unterschiedlichen Arten assoziativer Verknüpfungen in einem riesigen Netzwerk verschiedener Typen von Vorstellungen zu repräsentieren. Die Aktivierungsausbreitung in der Assoziationsmaschine geschieht automatisch, aber wir (System 2) besitzen in gewissem Umfang die Fähigkeit, die Gedächtnissuche zu steuern und sie überdies so zu programmieren, dass die Wahrnehmung eines Ereignisses in der Umwelt die Aufmerksamkeit darauf lenken kann. Als Nächstes wollen wir uns eingehender mit den Stärken und Beschränkungen von System 1 befassen.

## Normalität beurteilen

Die hauptsächliche Funktion von System 1 besteht darin, ein Modell unserer persönlichen Welt, in dem das repräsentiert ist, was normal in unserer Welt ist, aufrechtzuerhalten und zu aktualisieren. Das Modell ist aufgebaut auf Assoziationen, die Vorstellungen mit Situationen, Ereignissen, Handlungen und Ergebnissen verknüpfen, die mit einer gewissen Regelmäßigkeit gemeinsam auftreten, entweder gleichzeitig oder innerhalb einer relativ kurzen Zeitspanne. In dem Maße, wie sich diese Verknüpfungen bilden und verstärken, entsteht ein Netzwerk assoziierter Vorstellungen, das die Struktur von Ereignissen in unserem Leben repräsentiert, und es bestimmt unsere Interpretation der Gegenwart sowie unsere Zukunftserwartungen.

Die Fähigkeit zur Überraschung ist ein wesentlicher Aspekt unseres geistigen Lebens, und Überraschung selbst ist ein empfindlicher Indikator unseres Weltverständnisses und unserer Erwartungen an die Welt. Es gibt zwei wesentliche Spielarten der Überraschung. Einige Erwartungen sind aktiv und bewusst – Sie wissen, dass Sie darauf warten, dass ein bestimmtes Ereignis eintritt. Wenn die Stunde naht, erwarten Sie vielleicht, das Geräusch der Tür zu hören, wenn Ihr Kind aus der Schule kommt; wenn die Tür aufgeht, erwarten Sie, eine vertraute Stimme zu hören. Sie sind überrascht, wenn ein aktiv erwartetes Ereignis nicht eintritt. Viel größer dagegen ist die Kategorie der Ereignisse, die Sie passiv erwarten; Sie warten nicht auf sie, aber Sie sind nicht überrascht, wenn sie sich ereignen. Dies sind Ereignisse, die in einer Situation normal, aber nicht hinreichend wahrscheinlich sind, um aktiv erwartet zu werden.

Eine einzige Begebenheit mag eine Wiederholung weniger überraschend machen. Vor einigen Jahren machten meine Frau und ich in einer kleinen Ferienanlage auf einer Insel des Großen Barriereriffs Urlaub. Auf der Insel befinden sich nur vierzig Fremdenzimmer. Als wir zum Abendessen kamen, waren wir überrascht, als wir einen Bekannten trafen, einen Psychologen namens Jon. Wir begrüßten uns herzlich und äußerten unser Erstaunen über diesen Zufall. Jon reiste am nächsten Tag ab. Etwa zwei Wochen später waren wir in einem Theater in London. Ein Nachzügler setzte sich neben mich, nachdem die Lichter ausgegangen waren. Als in der Pause die Lichter

wieder angingen, sah ich, dass mein Nachbar Jon war. Meine Frau und ich tauschten uns später darüber aus, dass uns gleichzeitig zwei Gedanken durch den Kopf gegangen waren: erstens, dass dieses Zusammentreffen ein bemerkenswerterer Zufall war als die erste Begegnung; zum anderen waren wir beim zweiten Mal erheblich *weniger* überrascht als beim ersten Mal. Offenbar hatte die erste Begegnung irgendwie die mentale Repräsentation von Jon in uns verändert. Er war jetzt »der Psychologe, der auftaucht, wenn wir ins Ausland reisen«. Wir (System 2) wussten, dass dies eine aberwitzige Vorstellung war, aber unser System 1 hatte dafür gesorgt, dass es uns fast normal vorkam, Jon an exotischen Orten zu begegnen. Es hätte uns viel mehr überrascht, wenn ein anderer Bekannter als Jon in einem Theater in London neben uns gesessen hätte. Nach allen Wahrscheinlichkeitsmaßstäben war es viel unwahrscheinlicher, Jon in dem Theater zu begegnen als irgendeinem anderen unserer etlichen Hundert Bekannten – und doch wirkte es normaler, Jon zu treffen.

Unter bestimmten Bedingungen werden passive Erwartungen sehr schnell aktiv, wie wir bei einem anderen Zufall herausfanden. Vor einigen Jahren fuhren wir an einem Sonntagabend von New York City nach Princeton, wie wir es seit langer Zeit jede Woche taten. Unterwegs bot sich uns ein ungewohnter Anblick: Am Straßenrand brannte eine Auto. Als wir am Sonntag darauf an derselben Stelle vorbeikamen, brannte dort wieder ein Auto. Hier stellten wir wieder fest, dass wir beim zweiten Mal deutlich weniger überrascht waren als beim ersten Mal. Dies war jetzt »die Stelle, wo Autos in Flammen aufgehen«. Da die Umstände der Wiederholung die gleichen waren, genügte der zweite Vorfall, um eine aktive Erwartung zu erzeugen: Noch Monate, vielleicht sogar Jahre nach dem Ereignis dachten wir jedes Mal, wenn wir an dieser Stelle vorbeikamen, unwillkürlich an brennende Autos, und wir erwarteten, ein weiteres zu sehen (aber das geschah selbstverständlich nicht).

Der Psychologe Dale Miller und ich schrieben einen Aufsatz, in dem wir zu erklären versuchten, wie es dazu kommt, dass Ereignisse als normal oder abnormal wahrgenommen werden. Ich werde ein Beispiel aus unserer Beschreibung der Normentheorie verwenden, obwohl sich meine Interpretation leicht verändert hat:

> Einem Beobachter, der beiläufig die Gäste an einem Nachbartisch in einem modischen Restaurant beobachtet, fällt auf, dass der erste Gast, der die Suppe probiert, wie vor Schmerzen zusammenzuckt. Dieser Vorfall wird die Normalität einer Vielzahl von Ereignissen verändern. Es ist nun keine Überraschung mehr, wenn der Gast, der die Suppe zuerst probierte, heftig zusammenfährt, wenn er von einem Ober berührt wird; es ist auch keine Überraschung mehr, wenn ein anderer Gast einen Schrei unterdrückt, wenn er Suppe aus derselben Schüssel probiert. Diese Ereignisse und viele andere erscheinen jetzt normaler, als es sonst der Fall gewesen wäre, aber nicht unbedingt, weil sie vorgängigen Erwartungen entsprechen. Vielmehr erscheinen sie normal, weil sie die ursprüngliche Episode aktualisieren, aus dem Gedächtnis abrufen und im Zusammenhang mit ihr interpretiert werden.[1]

Stellen Sie sich vor, Sie selbst wären der Beobachter im Restaurant. Die ungewöhnliche Reaktion des ersten Gastes auf die Suppe würde Sie überraschen, und die Schreckreaktion auf die Berührung durch den Ober würde Sie ebenfalls überraschen. Doch das zweite ungewöhnliche Ereignis wird das erste ins Gedächtnis rufen, und beide zusammen ergeben Sinn. Die beiden Ereignisse fügen sich in ein Muster, wonach der Gast ein ungewöhnlich stark angespannter Mensch ist. Wenn andererseits nach dem ersten Gast ein zweiter, das Gesicht verziehend, die Suppe zurückgehen lässt, werden die beiden Überraschungen miteinander verknüpft, sodass jetzt höchstwahrscheinlich der Suppe die Schuld gegeben wird.

»Wie viele Tiere jeder Art nahm Moses mit in die Arche?« Die Zahl der Personen, denen auffällt, was an dieser Frage nicht stimmt, ist so gering, dass sie »Moses-Illusion« genannt wird. Moses nahm keine Tiere mit in die Arche; Noah tat dies. Wie der Vorfall mit dem zusammenzuckenden Suppenesser lässt sich auch die Moses-Illusion mühelos mit der Normentheorie erklären. Die Vorstellung, dass Tiere in die Arche gehen, ruft einen biblischen Kontext auf, und Moses ist in diesem Kontext keine ungewöhnliche Figur. Man erwartet ihn nicht unbedingt, aber die

Erwähnung seines Namens ist keine Überraschung. Auch die Tatsache, dass Moses und Noah den gleichen Vokallaut und die gleiche Silbenzahl haben, erleichtert die Verwechslung. Wie bei Worttriaden, die kognitive Leichtigkeit auslösen, registriert man unbewusst eine assoziative Kohärenz zwischen »Moses« und »Arche« und akzeptiert die Frage daher sehr schnell. Ersetzt man in diesem Satz »Moses« durch »George W. Bush«, hat man einen schlechten politischen Witz, aber keine Illusion.

Wenn etwas Zement nicht in den gegenwärtigen Kontext aktivierter Vorstellungen passt, erkennt das System eine Anomalie – wie Sie es gerade erlebt haben. Sie hatten keine konkrete Vorstellung davon, was nach »etwas« kommen sollte, aber als das Wort »Zement« erschien, wussten Sie, dass es in diesem Satz anomal war. Neuropsychologische Hirnstudien haben gezeigt, dass Anomalien erstaunlich schnell und scharfsinnig erkannt werden. In einem neueren Experiment hörten Versuchspersonen den Satz: »Die Erde umläuft jedes Jahr das Problem.« In der Hirnaktivität tauchte innerhalb von zwei Zehntelsekunden nach Beginn der Darbietung des ungewöhnlichen Wortes ein spezifisches Muster auf. Noch bemerkenswerter ist die Tatsache, dass die gleiche Hirnreaktion genauso schnell eintritt, wenn eine männliche Stimme sagt: »Ich glaube, ich bin schwanger, weil mir jeden Morgen übel ist.« Oder wenn die Stimme einer Person, die Upperclass-Englisch spricht, sagt: »Ich hab ein großes Tattoo auf meinem Rücken.«[2] Eine riesige Menge an Weltwissen muss unverzüglich aktiviert werden, damit die Inkongruenz erkannt wird: Die Stimme muss als Upperclass-Englisch identifiziert werden, und diese Tatsache muss mit der allgemeinen Feststellung, dass große Tattoos in der Oberschicht unüblich sind, kurzgeschlossen werden.

Wir können miteinander kommunizieren, weil unser Weltwissen und unser Sprachgebrauch weitgehend übereinstimmen. Wenn ich einen Tisch erwähne, ohne ihn genauer zu beschreiben, verstehen Sie, dass ich einen normalen Tisch meine. Sie wissen mit Sicherheit, dass seine Oberfläche annähernd eben ist und dass er deutlich weniger als 25 Füße hat. Wir besitzen Normen für eine immense Zahl von Kategorien, und diese Normen bilden den Hintergrund für das sofortige Erkennen von Anomalien wie schwangeren Männern und tätowierten Aristokraten.

Um die Rolle von Normen in der Kommunikation zu ermessen, betrachten Sie den folgenden Satz: »Die große Maus kletterte über den Rüssel des sehr kleinen Elefanten.« Ich kann davon ausgehen, dass Ihre

Normen für die Größe von Mäusen und Elefanten sich nicht allzu sehr von meinen unterscheiden. Die Normen spezifizieren eine typische oder durchschnittliche Größe für diese Tiere, und sie enthalten auch Informationen über die Schwankungsbreite innerhalb der Kategorie. Es ist sehr unwahrscheinlich, dass einer von uns beiden vor seinem geistigen Auge sieht, wie eine Maus, die größer ist als ein Elefant, über einen Elefanten hinwegschreitet, der kleiner ist als eine Maus. Vielmehr stellten wir uns getrennt, aber gemeinsam eine Maus vor, die kleiner als ein Schuh ist und über einen Elefanten klettert, der größer als ein Sofa ist. System 1, das Sprache versteht, hat Zugriff auf Normen von Kategorien, die die Spannweite plausibler Werte sowie die typischsten Fälle spezifizieren.

## Ursachen und Intentionen

»Freds Eltern trafen verspätet ein. Der Partyservice wurde in Kürze erwartet. Fred war wütend.« Sie wissen, weshalb Fred wütend war, und er war es nicht etwa deshalb, weil der Partyservice in Kürze erwartet wurde. In Ihrem Netzwerk von Assoziationen sind Wut und mangelnde Pünktlichkeit als Wirkung und deren mögliche Ursache miteinander verknüpft, aber zwischen Wut und der Vorstellung, den Partyservice zu erwarten, besteht kein derartiger Zusammenhang. Während Sie lasen, konstruierten Sie sofort eine kohärente Geschichte; Sie kannten sofort die Ursache von Freds Wut. Das Auffinden solcher kausalen Verknüpfungen ist ein Teil des Verstehens einer Geschichte und eine automatische Operation von System 1. System 2, Ihrem bewussten Selbst, wurde die beiläufige Interpretation angeboten, und es akzeptierte sie.

Eine Geschichte in Nassim Talebs Buch *Der Schwarze Schwan* veranschaulicht diese automatische Suche nach kausalen Verknüpfungen. Er berichtet, die Kurse von Anleihen seien an dem Tag gestiegen, an dem Saddam Hussein in seinem Versteck im Irak aufgespürt worden sei. Offenbar suchten die Investoren an jenem Morgen sicherere Anlagen, und Bloomberg News brachte die folgende Schlagzeile: »Kurse von US-Schatzwechseln steigen; Festnahme von Hussein wird den Terrorismus vermutlich nicht eindämmen.« Eine halbe Stunde später sanken die Anleihekurse wieder, und die korrigierte Schlagzeile lautete: »Kurse von US-Schatzwechseln sinken; Festnahme von Hussein steigert die Anziehungskraft risikobehafteter Anlagen.« Offensichtlich war Husseins Festnahme das wichtigste Ereignis des Tages, und weil die automatische Suche nach Ursachen unser Denken maßgeblich beeinflusst, war dieses Ereignis die bevorzugte Erklärung für alles, was an jenem Tag am Markt geschah. Die beiden Schlagzeilen wirken auf den ersten Blick wie Erklärungen der Vorgänge auf dem Markt, aber eine Aussage, die zwei gegensätzliche Ergebnisse erklären kann, erklärt gar nichts. Tatsächlich tun Schlagzeilen nichts anderes, als unser Bedürfnis nach Kohärenz zu befriedigen: Man nimmt an, dass ein bedeutendes Ereignis Folgen hat, und Folgen brauchen als Erklärung Ursachen. Wir verfügen nur über begrenzte Informationen über tagesaktuelle Ereignisse, und System 1 ist geschickt

darin, eine kohärente kausale Geschichte zu konstruieren, welche die ihm zur Verfügung stehenden Wissensfragmente miteinander verknüpft.

Lesen Sie den folgenden Satz:

> Nachdem Jane einen Tag damit verbracht hatte, in den überfüllten Straßen von New York hübsche Aussichten zu erkunden, bemerkte sie, dass ihr Geldbeutel verschwunden war.

Als Personen, die diese kurze Geschichte (zusammen mit vielen anderen) gelesen hatten, einem überraschenden Gedächtnistest unterzogen wurden, assoziierten sie das Wort »Taschendieb« häufiger mit der Geschichte als das Wort »Aussichten«, obwohl Letzteres im Satz vorkam, während Ersteres dies nicht tat.[3] Die Regeln assoziativer Kohärenz sagten uns, was geschehen war. Das Ereignis eines verlorenen Geldbeutels könnte viele verschiedene Ursachen evozieren: Der Geldbeutel rutschte aus einer Tasche heraus, wurde in einem Restaurant liegen gelassen usw. Aber wenn die Vorstellungen »verlorener Geldbeutel«, »New York« und »Menschenmenge« miteinander verbunden sind, rufen sie gemeinsam die Erklärung hervor, dass ein Taschendieb für den Verlust verantwortlich ist. In der Geschichte von der erschreckenden Suppe bringt das Ergebnis – ob das Zusammenzucken eines zweiten Gastes beim Probieren der Suppe oder die extreme Reaktion der ersten Person auf die Berührung durch den Ober – eine assoziativ kohärente Interpretation der anfänglichen Überraschung hervor, die eine plausible Geschichte vervollständigt.

Der aristokratische belgische Psychologe Albert Michotte veröffentlichte 1945 ein Buch, das jahrhundertealte Kausalitätskonzeptionen, die zumindest bis auf Humes Studie über die Mechanismen der Ideenassoziation zurückgingen, über den Haufen warf. Nach gängiger Auffassung schlossen wir aus wiederholten Beobachtungen von Korrelationen zwischen Ereignissen auf physikalische Kausalität. Wir hatten unzählige Erlebnisse, bei denen wir sahen, wie ein bewegter Gegenstand einen anderen berührte, der sofort anfing, sich zu bewegen, oftmals (aber nicht immer) in die gleiche Richtung. Dies geschieht, wenn eine Billardkugel gegen eine andere stößt, und das geschieht auch, wenn Sie

eine Vase umwerfen. Michotte hatte einen anderen Ansatz: Er behauptete, dass wir Kausalität genauso direkt *sehen*, wie wir Farbe sehen. Um diese Hypothese zu beweisen, führte er ein Experiment durch, bei dem die Versuchspersonen ein gezeichnetes schwarzes Quadrat in Bewegung sehen; es berührt ein anderes Quadrat, das sich sofort zu bewegen beginnt. Die Beobachter wissen, dass es keinen realen physischen Kontakt gibt, trotzdem haben sie eine starke »Kausalitätsillusion«. Wenn sich das zweite Objekt sofort zu bewegen beginnt, sagen sie, es sei vom ersten »angestoßen« worden. Experimente haben gezeigt, dass sechs Monate alte Säuglinge die Folge von Ereignissen als ein Ursache-Wirkung-Szenario sehen, und sie lassen Überraschung erkennen, wenn die Folge verändert wird.[4] Wir sind offensichtlich von Geburt an darauf eingestellt, *Eindrücke* von Kausalität zu haben, die nicht davon abhängen, ob wir über Muster der Verursachung nachdenken. Sie sind Produkte von System 1.

Im Jahr 1944, etwa zur selben Zeit, als Michotte seine Versuche über physikalische Kausalität veröffentlichte, benutzten die Psychologen Fritz Heider und Mary-Ann Simmel eine ähnliche Methode wie Michotte, um die Wahrnehmung *intentionaler* Kausalität zu untersuchen. Sie stellten einen Film her, der nur eine Minute und vierzig Sekunden lang ist und in dem man ein großes Dreieck, ein kleines Dreieck und einen Kreis sieht, die sich um eine Figur bewegen, die aussieht wie die schematische Ansicht eines Hauses mit einer offenen Tür. Die Betrachter sehen ein aggressives großes Dreieck, das ein kleineres Dreieck drangsaliert, einen verschreckten Kreis, wobei sich der Kreis und das kleine Dreieck verbünden, um den Rüpel zu überwältigen; sie sehen auch viel Gezerre an einer Tür und dann ein explosives Finale.[5] Der Eindruck von Absicht und Emotionalität ist unwiderstehlich; nur Menschen, die an Autismus leiden, erleben dies nicht. Natürlich geschieht dies nur in unserem Kopf. Unser Gehirn ist nicht nur bereit, sondern regelrecht darauf aus, Akteure zu identifizieren, ihnen Persönlichkeitszüge und spezifische Intentionen zuzuschreiben und ihre Handlungen als Ausdruck individueller Neigungen zu interpretieren. Auch hier sprechen die empirischen Befunde dafür, dass wir mit einer Anlage für intentionale Attributionen geboren werden: Schon Säuglinge unter einem Jahr identifizieren Rüpel und Opfer und erwarten von einem Verfolger, dass er den kürzesten Weg nimmt, um das zu fangen, hinter dem er her ist.[6]

Die subjektive Erfahrung einer frei gewollten Handlung ist weitgehend losgelöst von der physikalischen Kausalität. Obgleich es Ihre Hand ist, die nach dem Salzstreuer greift, denken Sie über dieses Ereignis nicht in Begriffen physikalischer Kausalität nach. Sie erleben es als verursacht durch eine Entscheidung, die Ihr vom Körper losgelöstes *Selbst* traf, weil Sie ihr Essen salzen wollten. Viele Menschen beschreiben wie selbstverständlich ihre Seele als die Quelle und die Ursache ihrer Handlungen. Der Psychologe Paul Bloom stellte in einem Beitrag, der 2005 in der Zeitschrift *The Atlantic* veröffentlicht wurde, die provokative Behauptung auf, unsere angeborene Bereitschaft, physikalische und intentionale Kausalität zu trennen, erkläre die Tatsache, dass religiöse Überzeugungen praktisch universell seien. Er bemerkte: »Wir nehmen die Welt der Dinge als weitgehend getrennt von der Welt der Psyche wahr, und dies ermöglicht es uns, dass wir uns seelenlose Körper und körperlose Seelen vorstellen.« Die beiden Kausalitätsmodi, auf die unsere Wahrnehmung eingestellt ist, führen dazu, dass wir die beiden zentralen Glaubenssätze vieler Religionen wie selbstverständlich akzeptieren: Eine immaterielle Gottheit ist die letzte Ursache der materiellen Welt, und unsterbliche Seelen kontrollieren unsere Körper, solange wir leben, und verlassen diese wieder, sobald wir sterben.[7] Nach Blooms Auffassung wurden die beiden Kausalitätskonzepte jeweils gesondert von evolutionären Kräften ausgeformt, die die Ursprünge der Religion in die Struktur von System 1 einbauten.

Die Bedeutung kausaler Intuitionen ist ein wiederkehrendes Thema in diesem Buch, weil Menschen dazu neigen, kausales Denken unsachgemäß anzuwenden, nämlich auf Situationen, die statistisches Denken erfordern. Statistisches Denken leitet Schlüsse über Einzelfälle von Eigenschaften aus Kategorien und Gesamtheiten ab. Leider ist System 1 dieser Denkmodus nicht zugänglich; System 2 kann lernen, statistisch zu denken, aber nur wenige Menschen erhalten die notwendige Schulung.

Die Psychologie der Kausalität bildete die Grundlage für meine Entscheidung, psychische Prozesse mit Metaphern der Handlungskompetenz zu beschreiben, wobei ich mich wenig um Konsistenz bemühte. Manchmal stelle ich System 1 als einen Akteur mit gewissen Merkmalen und Präferenzen dar, manchmal als eine Assoziationsmaschine, die die Wirklichkeit durch ein komplexes Muster von Verknüpfungen repräsentiert. Das System und die Maschine sind Fiktionen; ich benutze sie

deshalb, weil sie unserem Kausalitätsdenken entsprechen. Heiders Dreiecke und Kreise sind keine realen Akteure – es drängt sich nur geradezu auf, sie als solche zu sehen. Es ist eine Frage der mentalen Ökonomie. Ich vermute, dass Sie es (wie ich) leichter finden, wenn wir die Psyche manchmal in Begriffen von Merkmalen und Intentionen (die beiden Systeme) und manchmal in Begriffen mechanischer Regelmäßigkeiten (die Assoziationsmaschine) beschreiben. Ich habe nicht vor, Sie davon zu überzeugen, dass die Systeme real sind, ebenso wenig wie Heider uns glauben machen wollte, dass das große Dreieck wirklich ein Rüpel ist.

## Zum Thema »Normen und Ursachen«

»Als sich zeigte, dass der zweite Bewerber ebenfalls ein alter Freund von mir ist, hat mich das nicht mehr so überrascht. Schon nach sehr wenigen Wiederholungen fühlt sich eine neue Erfahrung normal an!«

»Wenn wir die Reaktion auf diese Produkte untersuchen, sollten wir uns nicht ausschließlich auf den Durchschnittswert konzentrieren. Wir sollten das gesamte Spektrum normaler Reaktionen betrachten.«

»Sie kann sich nicht damit abfinden, dass sie einfach Pech hatte; sie muss unbedingt eine Ursache finden. Zu guter Letzt wird sie noch glauben, dass jemand ihre Arbeit absichtlich sabotiert.«

# *7. Eine Maschine für voreilige Schlussfolgerungen*

Seit meiner Jugend ist mir ein Spruch des großen Komikers Danny Kaye unvergesslich. Über eine Frau, die er nicht mochte, sagte er: »Sie ist am liebsten außer sich und mag es, voreilige Schlüsse zu ziehen.« Ich erinnere mich, dass ich diesen Spruch im ersten Gespräch mit Amos Tversky über die Rationalität statistischer Intuitionen erwähnte, und ich glaube heute, dass er eine passende Beschreibung der Funktionen von System 1 liefert. Voreilige Schlussfolgerungen sind dann effizient, wenn sie mit hoher Wahrscheinlichkeit zutreffen, wenn die Kosten eines gelegentlichen Fehlers akzeptabel sind und sie viel Zeit und Mühe sparen. Dagegen sind voreilige Schlüsse riskant, wenn die Situation unbekannt ist, viel auf dem Spiel steht und man keine Zeit hat, weitere Informationen zu sammeln. Unter solchen Umständen sind intuitive Fehler wahrscheinlich, die allerdings durch eine gezielte Intervention von System 2 verhindert werden können.

## Vernachlässigung von Ambiguität und Unterdrückung von Zweifeln

**Abbildung 6**

Was haben die drei Skizzen in Abbildung 6 gemeinsam? Die Antwort lautet, dass sie alle mehrdeutig sind. Sie haben den Inhalt des linken Kästchens zweifellos als A B C gelesen und den des rechten Kästchens als 12 13 14, aber die mittleren Elemente in beiden Kästchen sind gleich. Sie hätten die beiden Folgen genauso gut als A 13 C oder 12 B 14 lesen können, aber das haben Sie nicht getan. Warum nicht? Die gleiche Figur wird in einem Kontext von Buchstaben als Buchstabe gelesen und in einem Kontext von Zahlen als Zahl. Der gesamte Kontext beeinflusst die Interpretation jedes Elements. Die Gestalt ist nicht eindeutig, aber man springt zu einer Schlussfolgerung über ihre Identität, und man wird sich der Ambiguität, die beseitigt wurde, nicht bewusst.

Was Ann betrifft, stellten Sie sich vermutlich eine Frau vor, die an Geld dachte und auf ein Gebäude mit Geldautomaten und Tresoren zugeht. Aber diese plausible Interpretation ist nicht die einzig mögliche; der Satz ist mehrdeutig. Wenn ein früherer Satz »Sie trieben sanft den Fluss hinunter« gelautet hätte, hätten Sie sich eine völlig andere Szene vorgestellt. Wenn Sie gerade an einen Fluss dachten, ist das Wort »Bank« [i. S. von Sandbank] nicht mit Geld assoziiert. Da ein expliziter Kontext fehlte, erzeugte System 1 von sich aus einen wahrscheinlichen Kontext. Wir wissen, dass es System 1 ist, weil Sie sich der Wahl oder der Möglichkeit einer anderen Interpretation nicht bewusst waren. Wenn Sie nicht vor Kurzem Kanu gefahren sind, haben Sie vermutlich mehr Zeit damit verbracht, auf Banken zu gehen, als auf Flüssen zu treiben, und Sie haben diese Ambiguität entsprechend aufgelöst. Wenn System 1 unsicher ist, wettet es auf eine Antwort, und die Wetten werden von unseren bisherigen Erfahrungen bestimmt. Die Wettregeln sind intelligent: Jüngste Ereignisse und der

gegenwärtige Kontext haben das größte Gewicht bei der Interpretation. Wenn uns kein Ereignis aus jüngster Vergangenheit einfällt, kommen länger zurückliegende Erinnerungen zum Tragen. Zu Ihren frühesten und unvergesslichsten Erfahrungen gehört das Singen des Abc; Sie haben nicht »A13C« gesungen.

Der wichtigste Aspekt der beiden Beispiele besteht darin, dass eine klare Entscheidung getroffen wurde, aber Sie haben nichts davon mitbekommen. Nur eine Interpretation fiel Ihnen ein, und die Ambiguität wurde Ihnen nicht bewusst. System 1 protokolliert nicht die Alternativen, die es verwirft, oder auch nur die Tatsache, dass es Alternativen gab. Bewusste Zweifel gehören nicht zum Repertoire von System 1; dazu wäre es erforderlich, gleichzeitig an miteinander unvereinbare Interpretationen zu denken, wozu es mentaler Anstrengung bedürfte. Ungewissheit und Zweifel sind die Domäne von System 2.

## Die Vorliebe, Aussagen zu glauben und eigene Erwartungen zu bestätigen

Der Psychologe Daniel Gilbert, der weithin bekannt ist als Autor des Bestsellers *Ins Glück stolpern*, schrieb einmal einen Aufsatz mit dem Titel »Wie mentale Systeme glauben«. Darin entwickelte er eine Theorie des Glaubens und Nichtglaubens von Aussagen, die er auf den im 17. Jahrhundert lebenden Philosophen Baruch de Spinoza zurückführte. Gilbert behauptete, das Verstehen einer Aussage beginne mit dem Versuch, sie zu glauben: Man muss zuerst wissen, was die Aussage bedeuten würde, wenn sie wahr wäre. Erst dann kann man entscheiden, ob man sie *für unwahr hält* oder nicht. Der anfängliche Versuch, zu glauben, ist eine automatische Operation von System 1, die mit der Konstruktion der bestmöglichen Interpretation der Situation verbunden ist. Selbst eine unsinnige Aussage, so Gilbert, wird zunächst einmal geglaubt. Nehmen wir folgendes Beispiel: »Renken essen Bonbons.« Wahrscheinlich waren Ihnen vage Eindrücke von Fischen und Bonbons bewusst, während Ihr assoziatives Gedächtnis automatisch auf Verknüpfungen zwischen den beiden Vorstellungen durchmustert wurde, die eine sinnvolle Interpretation des unsinnigen Satzes erlauben würden.

Gilbert vertritt die Auffassung, dass das Nichtglauben eine Operation von System 2 ist, und er beschrieb ein elegantes Experiment, mit dem er das zu beweisen versuchte.[1] Den Versuchspersonen wurden unsinnige Aussagen dargeboten, wie etwa der Satz: »Ein Dinca ist eine Flamme«, worauf innerhalb weniger Sekunden ein einzelnes Wort, »wahr« oder »falsch«, folgte. Später wurden sie einem Gedächtnistest unterzogen, um herauszufinden, welche Sätze sie als »wahr« markiert hatten. In einer Variante des Experiments sollten sich die Probanden während der Aufgabe Ziffern einprägen. Die Störung von System 2 hatte eine selektive Wirkung: Sie erschwerte es den Probanden, falsche Sätze nicht zu glauben. Bei einem späteren Gedächtnistest kam heraus, dass die kognitiv erschöpften Probanden schließlich viele der falschen Sätze für wahr hielten. Die Lehre aus diesem Experiment ist bedeutsam: Wenn System 2 anderweitig beschäftigt ist, glauben wir fast alles. System 1 ist leichtgläubig und neigt

dazu, Aussagen für wahr zu halten; System 2 ist dafür zuständig, Aussagen anzuzweifeln und nicht zu glauben, aber System 2 ist manchmal beschäftigt und oft faul. Tatsächlich gibt es Hinweise dafür, dass sich Menschen eher von inhaltsleeren, überredenden Botschaften wie etwa Werbespots beeinflussen lassen, wenn sie müde und kognitiv erschöpft sind.

Die Operationen des assoziativen Gedächtnisses tragen zu einer allgemeinen Bestätigungstendenz bei. Auf die Frage »Ist Sam freundlich?« fallen Ihnen andere Episoden von Sams Verhalten ein, als wenn man Sie gefragt hätte: »Ist Sam unfreundlich?« Die gezielte Suche nach bestätigenden Hinweisen, »positive Teststrategie« genannt, ist auch die Methode, mit der System 2 eine Hypothese überprüft. Im Gegensatz zu den Regeln von Wissenschaftstheoretikern, die empfehlen, Hypothesen dadurch zu überprüfen, dass man sie zu widerlegen versucht, suchen Menschen (und recht häufig auch Wissenschaftler) eher nach Daten, die mit ihren gegenwärtigen Überzeugungen vereinbar sind. Die Bestätigungstendenz von System 1 begünstigt die unkritische Annahme von Vorschlägen und überzeichnet die Wahrscheinlichkeit extremer und unwahrscheinlicher Ereignisse. Wenn man Sie fragt, für wie wahrscheinlich Sie es halten, dass Kalifornien innerhalb der nächsten dreißig Jahre von einem Tsunami heimgesucht wird, sind die Bilder, die Ihnen durch den Sinn gehen, wahrscheinlich Bilder von Tsunamis, so wie es Gilbert für unsinnige Aussagen wie »Renken essen Bonbons« behauptete. Sie werden dazu neigen, die Wahrscheinlichkeit einer Katastrophe zu überschätzen.

## Überzogene emotionale Kohärenz – der Halo-Effekt

Wenn Sie die Politik des Präsidenten gut finden, dann gefallen Ihnen wahrscheinlich auch seine Stimme und sein Aussehen. Die Tendenz, alles – auch Dinge, die Sie gar nicht beobachtet haben – an einem Menschen zu mögen (oder nicht zu mögen), wird »Halo-Effekt« genannt. Der Terminus ist in der Psychologie seit hundert Jahren gebräuchlich, aber in der Alltagssprache hat er keine große Verbreitung gefunden. Das ist schade, denn der Halo-Effekt ist ein guter Name für eine weitverbreitete kognitive Verzerrung, die unsere Sicht von Menschen und Situationen maßgeblich beeinflusst. Er trägt mit dazu bei, dass die Repräsentation der Welt, die System 1 erzeugt, einfacher und kohärenter ist als die Wirklichkeit.

Sie lernen auf einer Party eine Frau namens Joan kennen und finden sie sympathisch und umgänglich. Bald darauf fragen Sie sich, ob man wegen einer karitativen Spende an sie herantreten könnte. Was wissen Sie über die Großzügigkeit Joans? Die richtige Antwort lautet, dass Sie praktisch nichts darüber wissen, weil wenig Grund zu der Annahme besteht, dass kontaktfreudige Menschen auch großzügige Spender für wohltätige Zwecke sind. Aber Joan ist Ihnen sympathisch, und wenn Sie an sie denken, erleben Sie unwillkürlich wieder dieses Gefühl der Zuneigung. Sie mögen auch Großzügigkeit und großzügige Menschen. Aufgrund dieser Assoziation neigen Sie nunmehr zu der Annahme, Joan sei großzügig. Und jetzt, wo Sie glauben, dass Joan großzügig ist, mögen Sie Joan vermutlich noch mehr als zuvor, weil Sie die Liste ihrer positiven Merkmale um Großzügigkeit erweitert haben.

Tatsächliche Beweise für Großzügigkeit fehlen in Joans Geschichte, und die Lücke wird durch eine Vermutung gefüllt, die der subjektiven emotionalen Reaktion auf sie entspricht. In anderen Situationen sammeln sich nach und nach Hinweise an, und die Interpretation wird von der Emotion bestimmt, die mit dem ersten Eindruck verbunden ist. In einem klassischen psychologischen Experiment, das nichts von seiner Relevanz eingebüßt hat, präsentierte Solomon Asch seinen Probanden Beschreibungen von zwei Menschen und bat sie, ihre Persönlichkeit zu beurteilen.[2] Was halten Sie von Alan und Ben?

Alan: intelligent – fleißig – impulsiv – kritisch – eigensinnig – neidisch
Ben: neidisch – eigensinnig – kritisch – impulsiv – fleißig – intelligent

Wenn Sie wie die meisten von uns sind, haben Sie eine viel günstigere Meinung von Alan als von Ben. Die ersten Merkmale in der Liste verändern die Bedeutung der später auftauchenden Merkmale. Der Eigensinn einer intelligenten Person wird vermutlich als gerechtfertigt beurteilt und ruft vielleicht sogar Respekt hervor, aber Intelligenz bei einer neidischen und eigensinnigen Person macht sie gefährlicher. Der Halo-Effekt ist auch ein Beispiel für unterdrückte Ambiguität: Wie das Substantiv »Bank« ist auch das Adjektiv »eigensinnig« doppeldeutig und wird in einer Weise interpretiert, die in den Kontext passt.

Es gibt viele Variationen dieses Forschungsthemas. Teilnehmer einer Studie sollten zunächst die ersten drei Adjektive betrachten, die Alan beschreiben; anschließend sollten sie die letzten drei Adjektive betrachten, die, wie ihnen gesagt wurde, eine andere Person charakterisieren sollten. Als sie sich die beiden Personen vorgestellt hatten, wurden die Teilnehmer gefragt, ob es plausibel sei, dass alle sechs Adjektive auf eine Person zutreffen, und die meisten hielten dies für unmöglich![3]

Die Abfolge, in der wir die Merkmale einer Person zur Kenntnis nehmen, ist oftmals vom Zufall abhängig. Aber die Abfolge ist wichtig, weil der Halo-Effekt die Bedeutung des ersten Eindrucks verstärkt, manchmal so weit, dass nachfolgende Informationen größtenteils unberücksichtigt bleiben. Zu Beginn meiner Laufbahn als Professor zensierte ich Prüfungsaufsätze von Studenten in der üblichen Weise. Ich nahm mir jeweils ein Testheft vor und las sämtliche Aufsätze dieses Studenten direkt nacheinander, wobei ich sie einen nach dem anderen zensierte. Dann berechnete ich die Gesamtnote und ging weiter zum nächsten Studenten. Schließlich fiel mir auf, dass meine Benotungen der Aufsätze in jedem Heft verblüffend einheitlich waren. Mich beschlich der Verdacht, dass meine Benotung von einem Halo-Effekt beeinflusst wurde und dass die erste Frage, die ich zensierte, sich unverhältnismäßig stark auf die Gesamtnote

auswirkte. Der Mechanismus war einfach: Wenn ich den ersten Aufsatz gut benotet hatte, gab ich dem Studenten einen Vertrauensbonus, wenn ich später auf eine vage oder zweideutige Formulierung stieß. Das schien vernünftig zu sein. Ein Student, der einen so guten ersten Aufsatz geschrieben hatte, würde doch im zweiten keinen dummen Fehler machen! Aber meine Vorgehensweise war höchst problematisch. Wenn ein Student zwei Arbeiten geschrieben hatte, von denen eine gut und eine schlecht war, käme ich, je nachdem, welche Arbeit ich zuerst las, zu unterschiedlichen Endnoten. Ich hatte den Studenten gesagt, die beiden Aufsätze zählten gleich, aber das stimmte nicht – der erste wirkte sich viel stärker auf die Endnote aus als der zweite. Das war unannehmbar.

Ich entschied mich für eine neue Vorgehensweise. Statt die Hefte nacheinander von Anfang bis Ende zu lesen, las und benotete ich zunächst die ersten Aufsätze aller Studenten und nahm mir dann die zweiten vor. Ich schrieb die Punktzahlen jeweils auf die Innenseite der Rückseite, damit ich nicht (nicht einmal unbewusst) voreingenommen wäre, wenn ich den zweiten Aufsatz las. Bald nachdem ich auf die neue Methode umgestellt hatte, machte ich eine befremdliche Entdeckung: Ich vertraute meiner Benotung jetzt viel weniger als früher. Der Grund dafür war, dass ich jetzt häufig ein Unbehagen verspürte, das ich bis dahin nicht gekannt hatte. Wenn ich enttäuscht war über den zweiten Aufsatz eines Studenten und die Rückseite des Heftes aufschlug, um eine schlechte Zensur einzutragen, fiel mir hin und wieder auf, dass ich den ersten Aufsatz dieses Studenten sehr gut bewertet hatte. Mir fiel auch auf, dass ich versucht war, die Diskrepanz zu verringern, indem ich die Note, die ich noch nicht hingeschrieben, änderte, und es fiel mir schwer, der einfachen Regel zu folgen, nie dieser Versuchung nachzugeben. Meine Zensuren für die Aufsätze eines bestimmten Studenten schwankten oftmals erheblich. Diese mangelnde Kohärenz verunsicherte und frustrierte mich.

Ich war jetzt weniger zufrieden und weniger überzeugt von meinen Zensuren als früher, aber ich erkannte, dass dies ein gutes Zeichen war, ein Hinweis darauf, dass die neue Vorgehensweise besser war. Die Konsistenz, die ich früher erlebt hatte, trog; sie erzeugte ein Gefühl kognitiver Leichtigkeit, und mein System 2 war froh, ohne weitere Anstrengung die Endnote zu akzeptieren. Indem ich mich bei der Bewertung der späteren Aufsätze stark von der ersten Zensur beeinflussen ließ, ersparte ich mir die Dissonanz, die mit der Erkenntnis verbunden gewesen wäre, dass ein und

derselbe Student einige Fragen sehr gut und andere schlecht beantwortete. Die unangenehme Inkonsistenz, die enthüllt wurde, als ich auf die neue Vorgehensweise umstellte, war real: Sie spiegelte sowohl die Tatsache wider, dass eine einzelne Frage den Wissensstand eines Studenten nicht hinreichend erfassen kann, als auch die Unzuverlässigkeit meiner eigenen Notengebung.

Die Vorgehensweise, die ich wählte, um den Halo-Effekt abzuschwächen, entspricht einem allgemeinen Grundsatz: Fehler dekorrelieren! Um zu verstehen, wie dieses Prinzip funktioniert, stellen Sie sich eine große Zahl von Beobachtern vor, denen Glasgefäße gezeigt werden, die Pennys enthalten, und die die Anzahl der Pennys in jedem Gefäß schätzen sollen. Wie James Surowiecki in seinem Bestseller *Die Weisheit der Vielen* schrieb, schneiden Einzelpersonen bei diesen Aufgaben sehr schlecht ab, während Pools individueller Urteile erstaunlich richtige Ergebnisse liefern.[4] Einige Personen schätzen die tatsächliche Zahl viel zu hoch, andere schätzen sie zu niedrig, aber wenn viele Urteile gemittelt werden, ist der Mittelwert im Allgemeinen recht nah an der Lösung. Der Mechanismus ist einfach: Alle Personen betrachten dasselbe Gefäß, und all ihre Urteile haben eine gemeinsame Basis. Andererseits sind die Fehler, die einzelne Personen machen, unabhängig von den Fehlern, die andere machen, und (sofern kein systematischer Fehler vorliegt) heben sich tendenziell gegenseitig auf. Aber die Magie der Fehlerreduktion funktioniert nur dann gut, wenn die Beobachtungen unabhängig voneinander sind und ihre Fehler nicht miteinander korrelieren. Wenn alle Beobachter den gleichen Urteilsfehler begehen, wird dieser durch die Anhäufung von Urteilen nicht vermindert. Wenn man den Beobachtern erlaubt, sich gegenseitig wirksam zu beeinflussen, verringert dies die Größe der Stichprobe und damit die Genauigkeit der Gruppenschätzung.

Um aus vielen verschiedenen Datenquellen die nützlichsten Informationen zu gewinnen, sollte man immer versuchen, diese Quellen unabhängig voneinander zu machen. Diese Regel gilt für jedes gute polizeiliche Ermittlungsverfahren. Wenn es mehrere Zeugen für eine Tat gibt, dürfen diese nicht miteinander sprechen, bevor sie ihre Aussage machen. Dies soll nicht nur geheimes Zusammenwirken gegnerischer Zeugen verhindern, es soll auch unvoreingenommene Zeugen davon abhalten, sich gegenseitig zu beeinflussen. Zeugen, die sich über Erlebnisse austauschen, machen bei ihren Aussagen ähnliche Fehler und vermindern so

den Gesamtwert der von ihnen gelieferten Informationen. Es ist immer gut, die Redundanz von Informationsquellen zu verringern.

Das Prinzip unabhängiger Urteile (und dekorrelierter Fehler) lässt sich unmittelbar anwenden, um Sitzungen effektiver zu gestalten – eine Aktivität, mit der Führungskräfte in Organisationen einen Großteil ihrer Arbeitstage verbringen. Schon eine einfache Regel hilft: Bevor ein Problem erörtert wird, sollten alle Ausschussmitglieder eine sehr kurze Zusammenfassung ihres Standpunktes zu Papier bringen. Dieses Verfahren sorgt dafür, dass die Vielfalt der Kenntnisse und Meinungen innerhalb der Gruppe optimal genutzt wird. Die gängige Praxis offener Diskussionen gibt den Meinungen derjenigen, die als Erste zu Wort kommen und ihren Standpunkt sehr bestimmt artikulieren und andere dadurch dazu veranlassen, sich hinter ihnen einzureihen, zu viel Gewicht.

## What you see is all there is

Eine meiner Lieblingserinnerungen aus den ersten Jahren meiner Zusammenarbeit mit Amos ist ein kurzer Sketch, den er immer wieder gern zum Besten gab. In einer perfekten Nachahmung eines Philosophie-Professors, bei dem er studierte, brummte Amos in Hebräisch mit einem starken deutschen Akzent: »Sie dürfen niemals den *Primat des Ist* vergessen.« Was genau sein Lehrer mit diesem Satz gemeint hatte, habe ich nie verstanden (und Amos wohl auch nicht), aber Amos' Witze hatten immer eine ernste Pointe. Dieser Satz fiel ihm (und schließlich auch mir) immer ein, wenn wir die bemerkenswerte Asymmetrie feststellten zwischen der Art und Weise, wie das Gehirn gegenwärtig verfügbare Informationen verarbeitet, und der, wie es mit Informationen umgeht, die es nicht hat.

Es ist ein zentrales Konstruktionsmerkmal der Assoziationsmaschine, dass sie nur aktivierte Vorstellungen repräsentiert. Informationen, die nicht (und sei es unbewusst) aus dem Gedächtnis abgerufen werden, existieren gewissermaßen nicht. System 1 versteht sich hervorragend darauf, die bestmögliche Geschichte zu konstruieren, die momentan aktivierte Vorstellungen einbezieht, aber es kann keine Informationen berücksichtigen, die es nicht hat.

Das Erfolgskriterium von System 1 ist die Kohärenz der Geschichte, die es erschafft. Die Menge und Qualität der Daten, auf denen die Geschichte beruht, ist weitgehend belanglos. Wenn Informationen knapp sind – was häufig der Fall ist –, fungiert System 1 als eine Maschine für »Urteilssprünge«. Betrachten wir folgendes Beispiel: »Wird Mindik eine gute Führungskraft sein? Sie ist intelligent und stark …« Ihnen fiel schnell eine Antwort ein, und diese lautete Ja. Sie wählten auf der Grundlage der sehr beschränkten Informationen, die verfügbar waren, die beste Antwort aus, aber Sie haben überstürzt geurteilt. Was, wenn die nächsten beiden Adjektive »korrupt« und »grausam« lauteten?

Beachten Sie, was Sie *nicht* getan haben, als Sie kurz an Mindik als eine Führungskraft dachten. Sie fragten sich nicht als Erstes: »Was müsste ich wissen, bevor ich mir eine Meinung über die Führungskompetenz einer Person bilde?« System 1 begann von dem ersten Adjektiv an, spontan zu arbeiten: intelligent ist gut, intelligent und stark ist sehr gut. Dies ist die

beste Geschichte, die sich aus zwei Adjektiven konstruieren lässt, und System 1 erstellte sie mit großer kognitiver Leichtigkeit. Die Geschichte wird in dem Maße korrigiert, wie neue Informationen verfügbar werden (wie etwa: »Mindik ist korrupt«), aber es gibt kein Warten und kein subjektives Unbehagen. Und es bleibt auch eine kognitive Verzerrung zugunsten des ersten Eindrucks. Die Kombination aus einem nach Kohärenz strebenden System 1 und einem trägen System 2 hat zur Folge, dass System 2 viele intuitive Überzeugungen unterstützt, in denen sich die von System 1 erzeugten Eindrücke recht genau widerspiegeln. Natürlich kann System 2 Daten auch systematischer und sorgfältiger verarbeiten und eine Liste von Punkten durchgehen, die überprüft werden müssen, bevor man eine Entscheidung trifft – denken Sie etwa an den Kauf eines Hauses, bei dem Sie gezielt nach Informationen suchen, die Sie nicht haben. Doch wird System 1 auch die sorgfältigeren Entscheidungen beeinflussen. Sein Input hört nie auf.

Voreilige Schlussfolgerungen (»Urteilssprünge«) auf beschränkter Datenbasis sind so wichtig für das Verständnis intuitiven Denkens und kommen in diesem Buch so häufig vor, dass ich eine etwas sperrige Abkürzung dafür verwenden werde: WYSIATI, was für *What you see is all there is* (in etwa: »Nur was man gerade weiß, zählt«) steht. System 1 ist völlig unempfindlich für die Qualität und Quantität der Informationen, aus denen Eindrücke und Intuitionen hervorgehen.

Amos berichtete zusammen mit zwei seiner Doktoranden an der Universität Stanford über eine Studie, die einen direkten Bezug zu WYSIATI hat; dabei ging es darum, die Reaktion von Menschen zu beobachten, die einseitige Informationen erhalten und dies auch wissen.[5] Den Teilnehmern wurden rechtliche Sachverhalte dargeboten, wie etwa der folgende:

> Am 3. September hielt sich der Kläger David Thornton, ein 43-jähriger Gewerkschaftsvertreter, zu einem Routinebesuch im Thrifty Drug Store Nr. 168 auf, um mit Gewerkschaftsmitgliedern zu sprechen. Zehn Minuten nach seinem Eintreffen stellte ihn der Geschäftsführer zur Rede und sagte ihm, er dürfe sich nicht länger in den Verkaufsräumen mit den gewerkschaftlich organisierten Mitarbeitern unterhalten.

> Vielmehr solle er sich in ihrer Pause in einem Hinterzimmer mit ihnen treffen. Eine solche Forderung ist nach dem Tarifvertrag mit Thrifty Drug zulässig, wurde aber bislang nie geltend gemacht. Als Mister Thornton protestierte, sagte man ihm, er habe die Wahl, entweder diesen Auflagen nachzukommen, den Laden zu verlassen oder verhaftet zu werden. Worauf Mister Thornton den Geschäftsführer darauf hinwies, er habe bisher immer bis zu zehn Minuten lang in den Verkaufsräumen mit den Mitarbeitern sprechen dürfen, solange der Geschäftsbetrieb dadurch nicht beeinträchtigt worden sei, und er würde sich eher festnehmen lassen, als das übliche Vorgehen bei seinen Routinebesuchen zu verändern. Daraufhin rief der Geschäftsführer die Polizei und ließ Mister Thornton wegen Hausfriedensbruchs in Handschellen legen. Nachdem seine Personalien aufgenommen und er für kurze Zeit in eine Arrestzelle gesperrt worden war, wurden alle Anschuldigungen gegen ihn fallen gelassen. Mister Thornton verklagt Thrifty Drug wegen widerrechtlicher Festnahme.

Zusätzlich zu diesen Hintergrundinformationen, die alle Probanden lasen, hielten Anwälte der beiden Parteien mündliche Vorträge vor verschiedenen Gruppen von Versuchspersonen. Der Anwalt des Gewerkschaftsvertreters beschrieb die Festnahme natürlich als einen Einschüchterungsversuch, während der Anwalt des Unternehmens behauptete, Gespräche in den Verkaufsräumen störten den Geschäftsbetrieb und der Geschäftsführer habe sich daher angemessen verhalten. Einige Probanden hörten, wie Geschworene, beide Seiten. Die Anwälte vermittelten keine nützlichen Informationen, die man nicht der Hintergrundgeschichte entnehmen konnte.

Die Probanden kannten das Versuchsdesign, und diejenigen, die die Argumente nur einer Seite hörten, hätten sich mühelos die Argumente der anderen Seite gedanklich vergegenwärtigen können. Trotzdem wirkte sich die Präsentation einseitiger Informationen sehr stark auf die Urteile aus. Außerdem waren sich Probanden, die nur einseitige Informationen erhielten, ihrer Urteile sehr viel sicherer als Probanden, die beide Seiten hörten. Genau das würde man erwarten, wenn die Stärke der inneren Überzeugung, die man erlebt, von der Kohärenz der Geschichte abhängig

ist, die man aus den verfügbaren Informationen konstruieren kann. Die Konsistenz der Informationen, nicht ihre Vollständigkeit, ist das, was für eine gute Geschichte maßgeblich ist. Tatsächlich ist es so: Es ist leichter, alles, was man weiß, in ein kohärentes Muster einzupassen, wenn man wenig weiß.

WYSIATI erleichtert es, jene Kohärenz und kognitive Leichtigkeit zu erreichen, die dafür sorgt, dass wir eine Aussage als wahr akzeptieren. Es erklärt, wieso wir schnell denken und in einer komplexen Welt partielle Informationen sinnvoll interpretieren können. Die kohärente Geschichte, die wir zusammenfügen, kommt der Wirklichkeit meistens so nahe, dass sie zielführende, adäquate Handlungen unterstützt. Doch lässt sich mit WYSIATI auch eine lange und heterogene Liste von Urteils- und Entscheidungsfehlern erklären, darunter die folgenden:

– Selbstüberschätzung: Wie aus der WYSIATI-Regel folgt, wirken sich weder die Quantität noch die Qualität der Informationen nennenswert auf den Grad der subjektiven Überzeugung aus. Das Ausmaß, in dem wir an unsere Überzeugungen glauben, hängt überwiegend von der Qualität der Geschichte ab, die wir über das erzählen können, was wir sehen, auch wenn wir nur wenig sehen. Wir berücksichtigen oftmals nicht die Möglichkeit, dass Informationen, die für unser Urteil maßgeblich sein sollten, nicht vorhanden sind – unser Wissen ist begrenzt. Außerdem neigt unser assoziatives System dazu, sich auf ein kohärentes Aktivierungsmuster einzustellen, und es unterdrückt Zweifel und Ambiguität.
– Framing-Effekte: Verschiedene Darbietungsweisen derselben Information rufen oftmals unterschiedliche Emotionen hervor. Die Aussage »Die Überlebenswahrscheinlichkeit liegt im ersten Monat nach der Operation bei 90 Prozent« ist beruhigender als die äquivalente Aussage »Die Sterblichkeit liegt innerhalb des ersten Monats nach der Operation bei 10 Prozent«. In ähnlicher Weise ist Aufschnitt, der als »90-prozentig fettfrei« beschrieben wird, anziehender als »Aufschnitt mit 10 Prozent Fett«. Die alternativen Formulierungen sind ganz offenkundig gleichbedeutend, aber eine Person sieht normalerweise nur eine Formulierung, und »nur was man sieht, zählt«.
– Basisratenfehler: Erinnern Sie sich an Steve, den sanftmütigen und ordentlichen Menschen, der für einen Bibliothekar gehalten wird. Die

Persönlichkeitsbeschreibung ist markant und anschaulich, und obgleich Sie bestimmt wissen, dass es mehr männliche Landwirte als Bibliothekare gibt, ist Ihnen diese statistische Tatsache höchstwahrscheinlich nicht in den Sinn gekommen, als Sie zum ersten Mal über die Frage nachdachten. Nur was Sie wussten, zählte.

## Zum Thema »Voreilige Schlussfolgerungen«

»Sie weiß nichts über die Führungsqualitäten dieser Person. Sie lässt sich allein von dem Halo-Effekt einer guten Präsentation leiten.«

»Wir wollen Fehler dekorrelieren, indem wir vor der Diskussion getrennte Urteile zu dem anstehenden Problem einholen. Unabhängige Bewertungen verschaffen uns mehr Information.«

»Sie trafen diese weitreichende Entscheidung auf der Grundlage des positiven Berichts eines Beraters. WYSIATI – *what you see is all there is*. Ihnen schien nicht bewusst zu sein, wie wenig Informationen sie hatten.«

»Sie wollten nicht mehr Informationen, die ihnen vielleicht ihre Geschichte verderben würde. WYSIATI.«

# *8. Wie wir Urteile bilden*

Die Zahl der Fragen, die wir beantworten können, ist unbegrenzt – ganz egal, ob es Fragen sind, die uns ein anderer stellt, oder Fragen, die wir uns selbst stellen. Und auch die Zahl der Merkmale, die man beurteilen kann, ist grenzenlos. Sie können die Anzahl der Großbuchstaben auf dieser Seite zählen, die Höhe der Fenster in Ihrem Haus mit der Höhe der Fenster in dem Haus auf der anderen Straßenseite vergleichen und die Wiederwahlchancen Ihres Abgeordneten auf einer Skala von »sehr gut« bis »katastrophal« einstufen. Die Fragen werden an System 2 gerichtet, das die Aufmerksamkeit darauf fokussieren und das Gedächtnis durchsuchen wird, um die Antworten zu finden. System 2 nimmt Fragen entgegen oder erzeugt sie – in beiden Fällen steuert es die Aufmerksamkeit und durchstöbert das Gedächtnis auf der Suche nach den Antworten. System 1 arbeitet anders. Es überwacht fortwährend, was außerhalb und innerhalb des Bewusstseins geschieht, und es beurteilt permanent verschiedene Aspekte der Situation, ohne bestimmte Absichten zu verfolgen und mit geringer oder gar keiner Anstrengung. Diese elementaren Bewertungen spielen eine wichtige Rolle bei intuitiven Urteilen, weil sie leicht durch schwierigere Fragen ersetzt werden können – genau darum geht es bei Heuristiken und kognitiven Verzerrungen. Zwei weitere Merkmale von System 1 unterstützen ebenfalls die Ersetzung eines Urteils durch ein anderes. Da ist zum einen die Fähigkeit, Werte in andere Dimensionen zu übersetzen, was man beispielsweise tut, wenn man eine Frage beantwortet, die den meisten Menschen leicht vorkommt: »Wenn Sam so groß wäre, wie er intelligent ist, wie groß wäre er dann?« Schließlich gibt es die mentale Schrotflinte. Die Absicht von System 2, eine spezifische Frage zu beantworten oder ein bestimmtes Merkmal der Situation zu bewerten, löst automatisch weitere Berechnungen aus, darunter auch elementare Bewertungen.

## Elementare Bewertungen

System 1 wurde von der Evolution so ausgeformt, dass es die Hauptprobleme, die ein Organismus lösen muss, um zu überleben, fortwährend bewertet. Wie ist die Lage? Gibt es eine Bedrohung oder eine größere Chance? Läuft alles normal? Sollte ich mich nähern, oder sollte ich ausweichen? Die Fragen besitzen für einen Menschen in einem städtischen Umfeld vielleicht nicht die gleiche Dringlichkeit wie für eine Gazelle in der Savanne, aber wir haben die neuralen Mechanismen geerbt, die sich evolutionär herausbildeten, um den Bedrohungsgrad fortlaufend zu bewerten, und diese Mechanismen wurden nicht abgeschaltet. Situationen werden unentwegt als positiv oder negativ bewertet und lösen entsprechend Flucht- oder Annäherungsreaktionen aus. Eine gute Stimmungslage und kognitive Leichtigkeit sind beim Menschen die Entsprechungen zu Bewertungen, die Sicherheit und Vertrautheit feststellen.

Nehmen wir als konkretes Beispiel einer elementaren Bewertung die Fähigkeit, auf einen Blick Freund und Feind zu unterscheiden. Dies erhöht die Überlebenschancen in einer gefährlichen Welt, und tatsächlich hat sich beim Menschen eine solche spezielle Fähigkeit entwickelt. Alex Todorov, mein Kollege an der Universität Princeton, hat die biologischen Ursprünge schneller Urteile darüber, wie sicher es ist, mit einem Fremden zu interagieren, erforscht. [1] Er zeigte, dass wir über die Fähigkeit verfügen, mit einem einzigen Blick ins Gesicht eines Fremden zwei potenziell entscheidende Tatsachen über diese Person in Erfahrung zu bringen: wie dominant (und damit potenziell bedrohlich) sie ist und wie vertrauenswürdig, ob sie also eher freundliche oder feindliche Absichten hegt.[2] Die Form des Gesichts liefert Anhaltspunkte zur Beurteilung der Dominanz: Ein »starkes«, eckiges Kinn ist ein solcher Hinweisreiz. Der Gesichtsausdruck (Lächeln oder Stirnrunzeln) liefert Hinweise für die Beurteilung der Absichten eines Fremden. Die Kombination eines eckigen Kinns mit heruntergezogenen Mundwinkeln mag nichts Gutes bedeuten.[3] Die Treffgenauigkeit des Lesens von Gesichtern ist weit davon entfernt, perfekt zu sein: Ein rundes Kinn ist kein zuverlässiger Indikator für Gutmütigkeit, und Lächeln kann (bis zu einem gewissen Grad) vorgetäuscht

sein. Trotzdem verschafft selbst eine unvollkommene Fähigkeit, Fremde zu beurteilen, einen Überlebensvorteil.

Dieser uralte Mechanismus findet in der modernen Welt eine neue Anwendung: Er übt einen gewissen Einfluss auf das Wahlverhalten aus. Todorov zeigte seinen Studenten Bilder von Männergesichtern, manchmal nur eine Zehntelsekunde lang, und bat sie, die Gesichter in Bezug auf mehrere Attribute, wie Liebenswürdigkeit und Kompetenz, zu beurteilen. Die Einschätzungen der Betrachter deckten sich weitgehend. Die Gesichter, die Todorov zeigte, waren keine zufällig ausgewählte Stichprobe: Es waren Wahlkampfporträts von Politikern, die um Ämter konkurrierten. Anschließend verglich Todorov die Wahlergebnisse mit den Kompetenzeinschätzungen, die die Princeton-Studenten auf Basis der kurzen Darbietung der Fotografien und ohne jeden politischen Kontext vorgenommen hatten. Bei etwa 70 Prozent der Wahlen von Senatoren, Kongressabgeordneten und Gouverneuren war der Wahlgewinner der Kandidat, dessen Gesicht die höchsten Kompetenzbewertungen erzielt hatte. Dieses bemerkenswerte Ergebnis wurde bald darauf bei den nationalen Wahlen in Finnland, bei den Wahlen zu den Bauaufsichtsausschüssen in England und bei verschiedenen Wahlkämpfen in Australien, Deutschland und Mexiko bestätigt. [4] Erstaunlicherweise (zumindest für mich) sagten in Todorovs Studie die Kompetenzeinschätzungen die Wahlergebnisse viel zuverlässiger vorher als die Sympathiewerte.

Todorov hat herausgefunden, dass wir bei der Beurteilung der Kompetenz die beiden Dimensionen der Stärke und Vertrauenswürdigkeit miteinander kombinieren. Gesichter, die Kompetenz ausstrahlen, verbinden ein starkes Kinn mit einem leichten, selbstbewussten Lächeln. Es gibt keine Belege dafür, dass diese Gesichtsmerkmale tatsächlich vorhersagen, wie gut Politiker in ihrem Amt sein werden. Aber Studien über die Reaktion des Gehirns auf Gewinner und Verlierer zeigen, dass wir biologisch prädisponiert sind, Kandidaten abzulehnen, denen die von uns hochgeschätzten Attribute fehlen – bei diesen Forschungen kam heraus, dass Verlierer stärkere (negative) emotionale Reaktionen hervorriefen. Dies ist ein Beispiel dessen, was ich in den folgenden Kapiteln eine »Urteilsheuristik« *( judgement heuristic)* nennen werde. Wähler versuchen sich einen Eindruck davon zu verschaffen, wie gut ein Kandidat sein Amt ausüben wird, und dabei greifen sie auf eine einfachere Bewertung zurück,

die schnell und automatisch erfolgt und die verfügbar ist, wenn System 2 seine Entscheidung treffen muss.

Politikwissenschaftler haben im Anschluss an Todorovs Forschungen eine Kategorie von Wählern identifiziert, für die die automatischen Präferenzen von System 1 mit besonders hoher Wahrscheinlichkeit eine große Rolle spielen. Sie fanden das, wonach sie suchten, bei politisch uninformierten Wählern, die sehr viel fernsehen. Wie erwartet, ist der Effekt der »vom Gesicht abgelesenen« Kompetenz auf das Abstimmungsverhalten bei schlecht informierten Wählern mit hohem Fernsehkonsum dreimal so stark wie bei besser informierten Personen, die weniger fernsehen.[5] Ganz offensichtlich ist die relative Bedeutung von System 1 für Wahlentscheidungen nicht für alle Wähler gleich. Wir werden weitere Beispiele für solche individuellen Unterschiede kennenlernen.

System 1 versteht natürlich Sprache, und das Verstehen basiert auf den elementaren Bewertungen, die routinemäßig im Rahmen der Wahrnehmung von Ereignissen und des Verstehens von Nachrichten ausgeführt werden. Zu diesen Bewertungen zählen Berechnungen der Ähnlichkeit und Repräsentativität, Kausalattributionen und Bewertungen der Verfügbarkeit von Assoziationen und Musterbeispielen. Sie werden auch dann ausgeführt, wenn es keinen spezifischen Satz von Regeln zur Aufgabenlösung *(task set)* gibt, obwohl die Ergebnisse dazu benutzt werden, die Aufgabenanforderungen in dem Maße zu erfüllen, wie sie auftreten.[6]

Die Liste elementarer Bewertungen ist lang, aber nicht jedes mögliche Attribut wird bewertet. Werfen Sie einen kurzen Blick auf Abbildung 7, die ein Beispiel dafür zeigt.

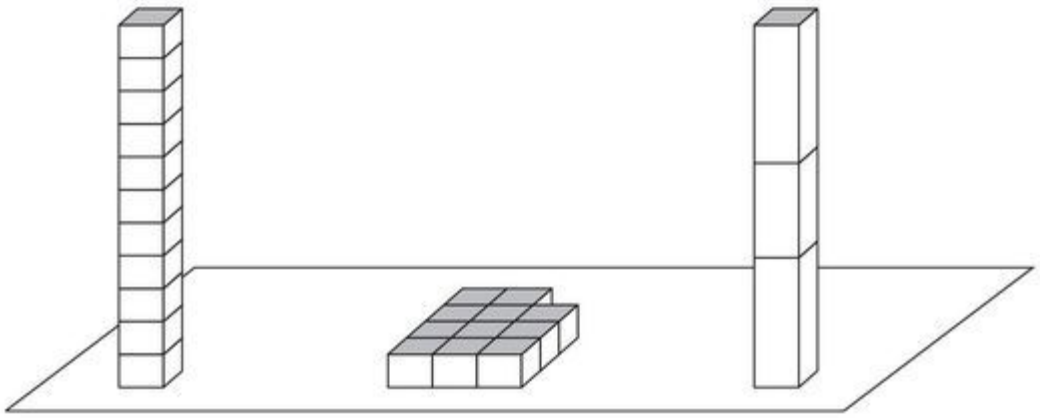

**Abbildung 7**

Ein flüchtiger Blick vermittelt einen sofortigen Eindruck von vielen Merkmalen des Schaubildes. Sie wissen, dass die beiden Türme gleich hoch sind und dass sie einander ähnlicher sind, als es der linke Turm und die Anordnung von Blöcken in der Mitte sind. Allerdings wissen Sie nicht sofort, dass die Anzahl der Blöcke im linken Turm gleich der Anzahl der Blöcke ist, die auf dem Boden angeordnet sind, und Sie haben keine Ahnung, wie hoch der Turm wäre, den Sie daraus errichten könnten. Um zu bestätigen, dass die Zahlen identisch sind, müssten Sie die beiden Mengen von Blöcken abzählen und die Ergebnisse vergleichen, etwas, das nur System 2 leisten kann.

## Mengen und Prototypen

Ein weiteres Beispiel ist die folgende Frage: Wie groß ist die durchschnittliche Länge der Geraden in Abbildung 8?

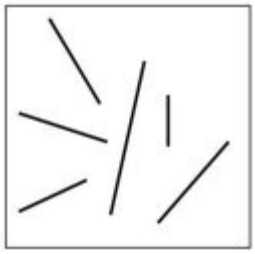

**Abbildung 8**

Die Frage ist leicht, und System 1 beantwortet sie ohne Weiteres. Experimente haben gezeigt, dass Menschen im Bruchteil einer Sekunde die durchschnittliche Länge einer Menge von Geraden mit großer Präzision erfassen können. Die Genauigkeit dieser Urteile wird auch nicht beeinträchtigt, wenn der Betrachter gedanklich von einer Gedächtnisaufgabe in Anspruch genommen wird. Sie können die durchschnittliche Länge nicht unbedingt in Zoll oder Zentimetern angeben, aber sie können die Länge einer anderen Linie so verändern, dass sie sehr gut mit dem Durchschnitt übereinstimmt. System 2 ist nicht erforderlich, um einen Eindruck von der durchschnittlichen Länge einer Menge von Geraden zu bilden. System 1 leistet dies automatisch und mühelos, genauso wie es die Farben der Linien und die Tatsache erkennt, dass sie nicht parallel sind. Wir können uns auch sofort einen Eindruck von der Anzahl der Objekte in einer Anordnung verschaffen – einen Eindruck, der präzise ist, wenn es höchstens vier Objekte sind, und nur ungefähr zutreffend, wenn es mehr sind.

Kommen wir jetzt zu einer anderen Frage: Was ist die Gesamtlänge der Geraden in Abbildung 8? Dies ist eine andere Sache, weil System 1 hier keine Vorschläge anzubieten hat. Diese Frage lässt sich nur dadurch beantworten, dass man System 2 aktiviert, das mühsam den Durchschnittswert schätzt, die Anzahl der Geraden schätzt oder abzählt und die durchschnittliche Länge mit der Zahl der Geraden multipliziert.

Die Unfähigkeit von System 1, auf einen Blick die Gesamtlänge einer Menge von Linien zu berechnen, mag Ihnen selbstverständlich erscheinen; Sie haben nie geglaubt, dass Sie das tun könnten. Tatsächlich ist es ein Beispiel für eine wichtige Begrenzung dieses Systems. Weil System 1 Kategorien durch einen Prototyp oder eine Menge typischer Exemplare repräsentiert, kommt es gut mit Durchschnittswerten, aber schlecht mit Summen zurecht. Die Größe der Kategorie, die Anzahl der Beispiele, die sie enthält, wird bei Urteilen über das, was ich »summenähnliche Variablen« nenne, tendenziell ignoriert.

Teilnehmer an einem der zahlreichen Experimente, die durch den Rechtsstreit infolge der verheerenden Ölpest, die der auf Grund gelaufene Tanker *Exxon Valdez* auslöste, angestoßen wurden, fragte man, ob sie bereit wären, für Netze zum Abdecken von Öltümpeln, in denen viele Zugvögel ertrinken, einen Preis zu bezahlen.[7] Verschiedene Probandengruppen erklärten ihre Bereitschaft, zu zahlen, um 2000, 20 000 oder 200 000 Vögel zu retten. Wenn die Rettung von Vögeln ein ökonomisches Gut ist, sollte es eine summenähnliche Variable sein: Die Rettung von 200 000 Vögeln sollte einem viel mehr wert sein als die Rettung von 2000 Vögeln. Tatsächlich beliefen sich die durchschnittlichen Beiträge der drei Gruppen auf jeweils 80, 78 und 88 Dollar. Die Zahl der Vögel machte keinen nennenswerten Unterschied. Die Teilnehmer in allen drei Gruppen reagierten auf einen Prototyp – das schreckliche Bild eines ertrinkenden Vogels, dessen Gefieder mit zähem Öl durchtränkt ist. Die fast vollständige Vernachlässigung der Quantität in solchen emotionalen Kontexten ist viele Male bestätigt worden.

## Intensitäten und wie man sie vergleichen kann

Fragen nach unserer Zufriedenheit, der Popularität des Präsidenten, der angemessenen Bestrafung von Wirtschaftskriminellen und den Zukunftsaussichten eines Politikers haben ein wichtiges Merkmal gemeinsam: Sie alle beziehen sich auf eine zugrunde liegende Dimension der Intensität oder eines Maßes, welche die Verwendung des Wortes »mehr« erlaubt: zufriedener, populärer, schwerer oder mächtiger (für einen Politiker). So kann die politische Zukunft einer Kandidatin von dem Tiefpunkt (»Sie wird bei den Vorwahlen ausscheiden«) bis zu dem Höhepunkt (»Sie wird eines Tages Präsidentin der Vereinigten Staaten werden«) reichen.

Hier begegnen wir einer neuen Fähigkeit von System 1. Eine grundlegende Intensitätsskala erlaubt eine Abstimmung *(matching)* über verschiedene Dimensionen hinweg. Wenn Verbrechen Farben wären, wäre Mord ein tieferer Rotton als Diebstahl. Wenn Verbrechen in Form von Musik ausgedrückt würden, würde Massenmord *fortissimo* gespielt werden, während die Anhäufung unbezahlter Strafzettel wegen Falschparkens nur ein schwaches *Pianissimo* wäre. Und natürlich haben wir ähnliche Gefühle hinsichtlich der Intensität von Strafen. In klassischen Experimenten passten Versuchspersonen die Lautstärke eines Tons der Schwere von Verbrechen an; andere Probanden passten die Lautstärke der Schwere der gesetzlichen Strafen an. Wenn Sie zwei Töne hören würden, einen für das Verbrechen und einen für die Strafe, würde es Ihnen ungerecht vorkommen, wenn ein Ton viel lauter als der andere wäre.[8]

Betrachten wir ein Beispiel, dem wir später erneut begegnen werden:

> Julie las flüssig, als sie vier Jahre alt war.

Stimmen wir jetzt Julies hervorragende Lesefähigkeit als Kind mit der folgenden Intensitätsskala ab:

Wie groß ist ein Mann, der so groß ist, wie Julie frühreif ist?

Was halten Sie von 1,80 Meter? Offenkundig zu klein. Wie wäre es mit 2,10 Meter? Vermutlich zu groß. Sie suchen nach einer Körpergröße, die genauso bemerkenswert ist wie die Leseleistung im Alter von vier Jahren. Ziemlich bemerkenswert, aber nicht außergewöhnlich. Mit 15 Monaten lesen zu können wäre außergewöhnlich, vielleicht wie ein Mann, der 2,30 Meter groß wäre.

Welcher Einkommenshöhe in Ihrem Beruf entspricht Julies Leseleistung?
Welches Verbrechen ist so schwer, wie Julie frühreif ist?
Welche Abschlussnote an einem Ivy-League-College (Eliteuniversitäten im Nordosten der USA) entspricht Julies Lesefähigkeit?

Nicht sehr schwer, oder? Außerdem können Sie davon ausgehen, dass Ihre Abgleiche weitgehend denen anderer Personen in Ihrem kulturellen Milieu entsprechen werden. Wenn Versuchspersonen aufgefordert werden, Julies Notendurchschnitt auf der Grundlage der Information über das Alter, in dem sie lesen lernte, vorherzusagen, beantworten sie dies dadurch, dass sie von einer Skala in eine andere Skala übersetzen und den passenden Notendurchschnitt auswählen. Wir werden sehen, wieso diese Methode der Vorhersage durch Abstimmung statistisch falsch ist – auch wenn sie für System 1 vollkommen natürlich ist, und für die meisten Menschen, mit Ausnahme von Statistikern, ist sie auch für System 2 akzeptabel.

## Die mentale Schrotflinte

System 1 führt zu jedem Zeitpunkt viele Berechnungen aus. Einige davon sind Routinebewertungen, die ständig ablaufen. Sobald Ihre Augen geöffnet sind, berechnet Ihr Gehirn eine dreidimensionale Repräsentation all dessen, was sich in Ihrem Gesichtsfeld befindet, einschließlich der Gestalt von Objekten, ihrer Positionen im Raum und ihrer Identitäten. Es bedarf keiner Absicht, um diese Operation oder die fortwährende Überwachung auf Verstöße gegen Erwartungen auszulösen. Im Unterschied zu diesen Routinebewertungen werden andere Berechnungen nur durchgeführt, wenn sie notwendig sind: Man bewertet nicht unentwegt, wie zufrieden oder vermögend man ist, und selbst wenn man politiksüchtig ist, bewertet man nicht in einem fort die Aussichten des Präsidenten. Die gelegentlichen Urteile erfordern eine gezielte Willensanstrengung. Man muss sie absichtlich herbeiführen.

Man zählt nicht automatisch die Anzahl der Silben jedes Wortes, das man liest, aber man kann es tun, wenn man sich dazu entscheidet. Doch unsere Kontrolle über intendierte Berechnungen ist keineswegs präzise: Wir berechnen oftmals viel mehr, als wir wollen oder brauchen. Ich nenne diese überschüssigen Berechnungen »mentale Schrotflinte«. Man kann mit einer Schrotflinte kein Punktziel anvisieren, weil sie Schrotkugeln verschießt, die streuen, und es scheint für System 1 fast genauso schwer zu sein, nicht mehr zu tun als das, was System 2 von ihm fordert. Zwei Experimente, deren Beschreibungen ich vor langer Zeit las, legten mir dieses Bild nahe.

Den Teilnehmern eines Experiments wurden Wortpaare vorgespielt, und sie sollten so schnell wie möglich eine Taste drücken, sobald sie bemerkten, dass sich die Wörter reimten.[9] Die Wörter in folgenden beiden Paaren reimen sich:

*Vote – Note*
*Vote – Goat*

Der Unterschied ist für Sie so offensichtlich, weil Sie die beiden Paare sehen. *Vote* und *Goat* reimen sich, aber sie werden unterschiedlich buchstabiert. Die Teilnehmer hörten nur die Wörter, aber sie wurden auch von der Schreibweise beeinflusst. Sie erkannten deutlich langsamer, dass sich die Wörter reimten, wenn deren Schreibweisen voneinander abwichen. Obgleich die Instruktionen von ihnen nur einen Vergleich der Laute verlangten, verglichen die Teilnehmer auch die Schreibung, und die Nichtübereinstimmung in der irrelevanten Dimension verlangsamte sie. Die Absicht, eine Frage zu beantworten, rief eine andere Frage hervor, die nicht nur überflüssig war, sondern die Lösung der Hauptaufgabe erschwerte.

In einer anderen Studie hörten die Versuchspersonen Satzfolgen, mit der Anweisung, so schnell wie möglich eine Taste zu drücken, um anzugeben, ob der Satz im wörtlichen Sinne wahr ist, und eine andere Taste, wenn der Satz im wörtlichen Sinne nicht wahr ist.[10] Wie lauten die richtigen Antworten für die folgenden Sätze?

Einige Straßen sind Schlangen.
Einige Arbeitsplätze sind Schlangen.
Einige Arbeitsplätze sind Gefängnisse.

Alle drei Sätze sind im wörtlichen Sinne falsch. Doch ist Ihnen vermutlich aufgefallen, dass der zweite Satz offenkundiger falsch ist als die anderen beiden – die bei dem Experiment gemessenen Reaktionszeiten bestätigten einen deutlichen Unterschied. Der Grund für diesen Unterschied liegt darin, dass die beiden schwierigen Sätze im übertragenen Sinne wahr sein können. Auch hier rief die Absicht, eine Berechnung durchzuführen, eine weitere Berechnung hervor. Aber auch hier setzte sich in dem Konflikt die richtige Antwort durch, auch wenn der Konflikt mit der irrelevanten Antwort die Leistung beeinträchtigte.

Im nächsten Kapitel werden wir sehen, dass die Kombination einer mentalen Schrotflinte mit Intensitätsabstimmungen erklärt, weshalb wir viele Dinge, über die wir wenig wissen, intuitiv beurteilen.

## Zum Thema »Urteile«

»Die Beurteilung der Attraktivität von Menschen ist eine elementare Bewertung. Wir tun dies automatisch, ob wir es wollen oder nicht, und es beeinflusst uns.«

»Es gibt Schaltkreise im Gehirn, die anhand der Gesichtsform die Dominanz der betreffenden Person einschätzen. Sie scheint für eine Führungsaufgabe geeignet zu sein.«

»Eine Strafe wird nur dann als gerecht empfunden, wenn ihre Intensität dem Verbrechen angemessen ist. So, wie man die Lautstärke eines Tons mit der Helligkeit eines Lichts zur Übereinstimmung bringen kann.«

»Dies war ein eindeutiger Fall einer mentalen Schrotflinte. Er wurde gefragt, ob er die Firma für finanziell solide halte, aber er konnte nicht vergessen, dass er ihre Produkte mochte.«

# *9. Eine leichtere Frage beantworten*

Ein bemerkenswerter Aspekt unseres geistigen Lebens besteht darin, dass wir nur selten überfragt sind. Es stimmt, gelegentlich sind wir mit einer Frage wie 17 × 24 = ? konfrontiert, deren Lösung uns nicht auf Anhieb einfällt, aber diese Momente der Ratlosigkeit sind selten. Unser Geist funktioniert normalerweise so, dass wir intuitive Gefühle und Meinungen über fast alles haben, was uns begegnet. Man findet Menschen sympathisch oder unsympathisch, lange bevor man viel über sie weiß; man vertraut oder misstraut Fremden, ohne zu wissen, wieso; man spürt, dass ein Unternehmen erfolgreich sein wird, ohne es zu analysieren. Unabhängig davon, ob wir sie explizit formulieren oder nicht, haben wir oft Antworten auf Fragen, die wir nicht vollständig verstehen, und wir stützen uns dabei auf Hinweise, die wir weder erklären noch verteidigen können.

## Fragen ersetzen

Ich schlage eine einfache Erklärung dafür vor, wie wir uns intuitive Meinungen über komplexe Sachverhalte bilden. Wenn eine befriedigende Antwort auf eine schwierige Frage nicht schnell gefunden wird, findet System 1 eine ähnliche Frage, die leichter ist, und beantwortet diese. Ich nenne diese Operation der Beantwortung einer Frage anstelle einer anderen »Ersetzung«. Außerdem führe ich die folgenden Bedingungen ein:

> Die Zielfrage liefert die Beurteilung, nach der man strebt.
> Die heuristische Frage ist die einfachere Frage, die man stattdessen beantwortet.

Eine Heuristik ist, technisch definiert, ein einfaches Verfahren, das uns hilft, adäquate, wenn auch oftmals unvollkommene Antworten auf schwierige Fragen zu finden. Das Wort hat die gleiche Wurzel wie der Ausruf »heureka!«. Schon zu einem frühen Zeitpunkt meiner Zusammenarbeit mit Amos entwickelten wir das Konzept der Ersetzung, das den Kern der späteren Theorie der Heuristiken und kognitiven Verzerrungen bildete. Wir fragten uns, wie es Menschen fertigbringen, Wahrscheinlichkeitsurteile abzugeben, ohne genau zu wissen, was Wahrscheinlichkeit bedeutet. Wir gelangten zu dem Schluss, dass Menschen diese unmögliche Aufgabe irgendwie vereinfachen müssen, und wir nahmen uns vor, herauszufinden, wie sie das tun. Unsere Antwort lautete, dass Menschen, wenn sie aufgefordert werden, die Wahrscheinlichkeit von etwas einzuschätzen, tatsächlich etwas anderes beurteilen, aber glauben, sie hätten die Wahrscheinlichkeit beurteilt. System 1 verfährt oftmals so, wenn es mit schwierigen Zielfragen konfrontiert ist, sofern es eine ähnliche, aber leichtere heuristische Frage müheloser beantworten kann.

Die Ersetzung einer Frage durch eine andere kann eine gute Strategie zur Lösung schwieriger Probleme sein, und George Pólya bezog die Ersetzungsmethode in sein klassisches Werk *Schule des Denkens* ein: »Wenn du ein Problem nicht lösen kannst, dann gibt es ein einfacheres

Problem, das du lösen kannst: Finde es.« Pólyas Heuristiken sind strategische Verfahren, die gezielt von System 2 umgesetzt werden. Die Heuristiken, die ich in diesem Kapitel diskutiere, werden nicht bewusst ausgewählt; vielmehr sind sie eine Folge der mentalen Schrotflinte, der Tatsache, dass wir nicht in der Lage sind, unsere Antworten zielgerichtet und präzise auf Fragen abzustimmen.

| ***Zielfrage*** | ***Heuristische Frage*** |
|---|---|
| Wie viel sind Sie bereit auszugeben, um eine bedrohte Art zu retten? Wie zufrieden sind Sie gegenwärtig mit Ihrem Leben? | Wie sehr berührt es mich, wenn ich an sterbende Delfine denke? Wie ist gerade meine Stimmung? |
| Wie beliebt wird der Präsident in sechs Monaten sein? | Wie beliebt ist der Präsident gegenwärtig? |
| Wie sollten Finanzberater, die ältere Menschen ausnehmen wollen, bestraft werden? | Wie viel Wut spüre ich, wenn ich an Finanzhaie denke? |
| Diese Frau kandidiert bei den Vorwahlen. Wie weit wird sie es in der Politik bringen? | Sieht diese Frau wie eine politische Gewinnerin aus? |

**Tabelle 1**

Betrachten Sie die Fragen, die in der linken Spalte von Tabelle 1 aufgelistet sind. Dies sind schwierige Fragen, und bevor Sie eine gut durchdachte Antwort auf eine davon geben können, müssen Sie andere schwierige Fragen lösen. Was bedeutet Zufriedenheit? Was sind die wahrscheinlichen politischen Entwicklungen in den kommenden sechs Monaten? Was sind die üblichen Strafen für Wirtschaftsverbrechen? Wie stark ist der Wettbewerb unter den Kandidaten? Welche anderen ökologischen oder sonstigen Ursachen sollten in Betracht gezogen werden?

Es wäre eine völlige Überforderung, all diese Fragen minutiös zu durchdenken, ehe man mit einer Antwort aufwarten kann. Aber wir sind nicht auf perfekt durchdachte Antworten auf Fragen beschränkt. Es gibt eine heuristische Alternative zu sorgfältigem Denken, die manchmal recht gut funktioniert und manchmal zu gravierenden Fehlern führt.

Die mentale Schrotflinte erleichtert es, schnelle Antworten auf schwierige Fragen zu finden, ohne unserem trägen System 2 allzu harte Arbeit aufzubürden. Mit hoher Wahrscheinlichkeit wird zu jeder der Fragen auf der linken Seite das rechte Pendant evoziert, das sehr leicht zu beantworten ist. Ihre emotionalen Einstellungen zu Delfinen und Finanzhaien, Ihre gegenwärtige Stimmung, Ihr Eindruck von den politischen Fähigkeiten der Vorwahl-Kandidatin oder die gegenwärtige Popularität des Präsidenten werden Ihnen mühelos einfallen. Die heuristischen Fragen liefern eine Standardantwort auf jede der schwierigen Zielfragen.

Etwas fehlt noch in dieser Geschichte: Die Antworten müssen auf die ursprünglichen Fragen zugeschnitten werden. So muss zum Beispiel meine gefühlsmäßige Reaktion auf sterbende Delfine in Dollars ausgedrückt werden. Eine weitere Fähigkeit von System 1, die Abstimmung der Intensität, kann dieses Problem lösen. Erinnern wir uns daran, dass sowohl Gefühle als auch Geldspenden Intensitätsskalen sind. Meine Gefühle für Delfine können mehr oder weniger stark sein, und es gibt einen Spendenbetrag, der der Intensität meiner Gefühle entspricht. Der Dollarbetrag, der mir spontan einfällt, ist der passende Betrag. Ähnliche Intensitätsentsprechungen gibt es für alle Fragen. So können zum Beispiel die politischen Fähigkeiten eines Kandidaten von »kläglich« bis »außerordentlich eindrucksvoll« reichen, und die Skala des politischen Erfolgs kann von dem Tief »Sie wird die Vorwahlen verlieren« bis zu dem Hoch »Sie wird eines Tages Präsidentin der USA werden« reichen.

Die automatischen Prozesse der mentalen Schrotflinte und der Intensitätsabstimmung stellen oftmals eine oder mehrere Antworten auf leichte Fragen bereit, die sich schwierigen Zielfragen zuordnen lassen. In einigen Fällen kommt es zur Ersetzung, und eine heuristische Antwort wird von System 2 unterstützt. Selbstverständlich hat System 2 die Gelegenheit, diese intuitive Antwort abzulehnen oder sie durch Einbeziehung weiterer Informationen zu modifizieren. Ein träges System 2 folgt allerdings oftmals dem Pfad des geringsten Widerstandes und unterstützt eine heuristische

Antwort, ohne genauer zu überprüfen, ob diese wirklich angemessen ist. Sie sind nicht überfragt, Sie müssen sich nicht sehr stark anstrengen, und Sie bemerken vielleicht nicht einmal, dass Sie nicht die Frage beantwortet haben, die Ihnen gestellt wurde. Außerdem wird Ihnen vielleicht nicht bewusst, dass die Zielfrage schwierig war, weil Ihnen so mühelos eine intuitive Antwort einfiel.[1]

## Die 3-D-Heuristik

Betrachten Sie das Bild mit den drei Männern, und beantworten Sie dann die nachstehende Frage.

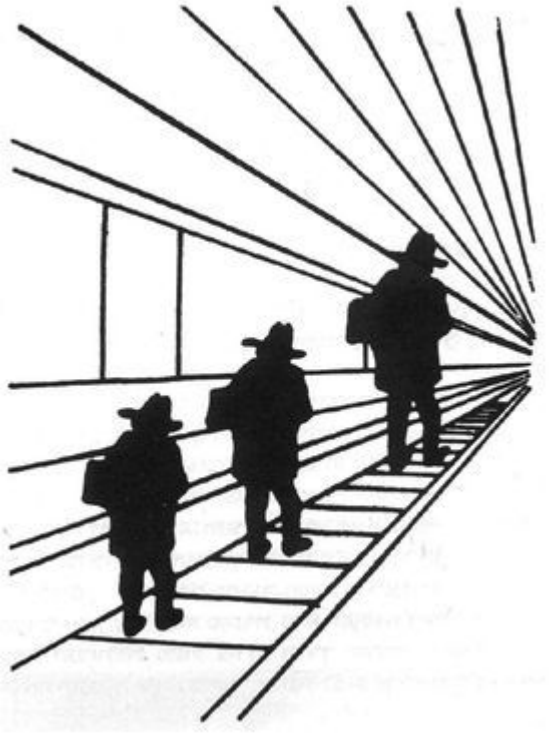

**Abbildung 9**

> Ist die Figur rechts, so wie sie auf dieser Seite gedruckt ist, größer als die Figur links?

Der Augenschein legt sofort eine Antwort nahe: Die rechte Figur ist größer. Doch wenn Sie beide Figuren mit einem Lineal vermessen, werden Sie feststellen, dass die Figuren genau gleich groß sind. Ihr Eindruck von der relativen Größe wird von einer mächtigen Illusion beherrscht, die den Ersetzungsprozess plastisch veranschaulicht.

Der Korridor, in dem sich die Figuren zu bewegen scheinen, ist perspektivisch dargestellt und scheint in die Tiefe zu führen. Ihr Wahrnehmungssystem interpretiert das Bild automatisch als eine dreidimensionale Szene, nicht als ein Bild, das auf eine flache Papieroberfläche gedruckt wurde. In der 3-D-Interpretation ist die Person rechts sowohl viel weiter weg als auch viel größer als die Person links. Für die meisten von uns ist der Eindruck der dreidimensionalen Größe überwältigend. Nur bildende Künstler und erfahrene Fotografen haben die Fähigkeit entwickelt, die Zeichnung als ein auf Papier gedrucktes Objekt zu sehen. Bei allen anderen vollzieht sich eine Ersetzung: Der dominante Eindruck der dreidimensionalen Größe diktiert das Urteil über die zweidimensionalen Größe. Diese Illusion ist auf eine 3-D-Heuristik zurückzuführen.

Was hier geschieht, ist eine echte Illusion, kein Missverstehen der Frage. Sie wussten, dass sich die Frage auf die Größe der Figuren in dem Bild bezogen hat, so wie es auf der Seite gedruckt ist. Wenn man Sie gebeten hätte, die Größe der Figuren abzuschätzen, wissen wir aus Experimenten, dass Sie die Größe in Zoll, nicht in Fuß angegeben hätten. Die Frage hat Sie nicht verwirrt, aber Sie wurden durch die Antwort auf eine Frage beeinflusst, die Ihnen nicht gestellt wurde: »Wie groß sind die drei Figuren?«

Der entscheidende Schritt bei der Heuristik – die Ersetzung der zweidurch die dreidimensionale Größe – ereignete sich automatisch. Das Bild enthält Hinweisreize, die eine dreidimensionale Interpretation nahelegen. Diese Hinweise sind für die anstehende Aufgabe irrelevant – die Beurteilung der Größe der Figur auf der Seite –, und Sie hätten sie ignorieren sollen, aber das konnten Sie nicht. Mit der Heuristik ist die Wahrnehmungsverzerrung verbunden, dass Objekte, die auf der Seite scheinbar weiter weg sind, auch größer erscheinen. Wie dieses Beispiel verdeutlicht, wird ein Urteil, das auf Ersetzung beruht, zwangsläufig in vorhersagbarer Weise verzerrt sein. In diesem Fall geschieht dies so tief in Ihrem Wahrnehmungssystem, dass Sie dem Eindruck nicht widerstehen können.

## Die Stimmungsheuristik für Glück

Eine Befragung deutscher Studenten liefert eines der besten Beispiele für Ersetzung. [2] Der Fragebogen, den die jungen Probanden ausfüllen sollten, enthielt die folgenden beiden Fragen:

> Wie glücklich fühlen Sie sich zurzeit?
> Wie viele Verabredungen hatten Sie im letzten Monat?

Die Experimentatoren interessierten sich für die Korrelation zwischen den beiden Antworten. Würden die Studenten, die angaben, viele Verabredungen gehabt zu haben, sagen, dass sie sich glücklicher fühlten als diejenigen mit weniger Verabredungen? Überraschenderweise war dies nicht der Fall: Die Korrelation zwischen den Antworten betrug ungefähr null. Offensichtlich waren Verabredungen nicht das, was den Studenten als Erstes einfiel, wenn sie aufgefordert wurden, ihr Glücksgefühl zu bewerten. Eine andere Studentengruppe sah die gleichen beiden Fragen, aber in umgekehrter Reihenfolge:

> Wie viele Verabredungen hatten Sie im letzten Monat?
> Wie glücklich fühlen Sie sich zurzeit?

Diesmal kamen ganz andere Ergebnisse heraus. In dieser Abfolge war die Korrelation zwischen der Anzahl der Verabredungen und dem angegebenen Glücksgefühl so hoch, wie Korrelationen zwischen psychologischen Messgrößen überhaupt nur sein können.[3] Was war geschehen?

Die Erklärung ist einfach, und sie ist ein gutes Beispiel für den Ersetzungsprozess. Verabredungen standen offenkundig nicht im Mittelpunkt des Lebens dieser Studenten (in der ersten Erhebungsgruppe waren Glücksgefühl und Verabredungen nicht miteinander korreliert), aber

als sie aufgefordert wurden, über ihr Liebesleben nachzudenken, reagierten sie eindeutig emotional darauf. Die Studenten, die viele Verabredungen hatten, wurden an einen glücklichen Aspekt ihres Lebens erinnert, während diejenigen, die keine Verabredungen hatten, an Einsamkeit und Zurückweisung erinnert wurden. Die durch die Verabredungsfrage hervorgerufene Emotion war bei allen noch präsent, als die Frage nach dem allgemeinen Glücksgefühl gestellt wurde.

Die psychischen Prozesse bei diesem Experiment entsprechen genau den psychischen Prozessen bei der Größenillusion in Abbildung 9. Die Frage nach dem »Glücksgefühl zurzeit« ist keine natürliche oder leichte Beurteilung. Eine gute Antwort erfordert gründliches Nachdenken. Doch die Studenten, die gerade nach ihren Verabredungen gefragt worden waren, mussten nicht angestrengt nachdenken, weil sie bereits eine Antwort auf eine ähnliche Frage präsent hatten: Wie glücklich sie mit ihrem Liebesleben seien. Sie ersetzten die Frage, die ihnen gestellt wurde, durch die Frage, auf die sie eine vorgefertigte Antwort hatten.

Wie bei der optischen Täuschung können wir auch hier fragen: Sind die Studenten nicht ganz klar bei Verstand? Glauben sie wirklich, dass die beiden Fragen – die Frage, die ihnen gestellt wurde, und die Frage, die sie beantworten – synonym sind? Natürlich nicht. Die Studenten verlieren nicht vorübergehend ihre Fähigkeit, das Liebesleben von dem Leben insgesamt zu unterscheiden. Nach den beiden Konzepten gefragt, würden sie sagen, dass man beides auseinanderhalten müsse. Aber sie wurden nicht danach gefragt, ob sich die Konzepte unterscheiden. Sie wurden gefragt, wie glücklich sie seien, und System 1 hatte eine Antwort parat.

Die Frage nach den Verabredungen ist allerdings nicht einzigartig. Das gleiche Muster findet man, wenn eine Frage nach den Beziehungen der Studenten zu ihren Eltern oder nach ihren Finanzen der Frage nach dem allgemeinen Glücksgefühl unmittelbar vorausgeht. In beiden Fällen war die Zufriedenheit in einem bestimmten Bereich ausschlaggebend für das berichtete Glücksgefühl. [4] Jede emotional bedeutsame Frage, die die Stimmung einer Person verändert, hat den gleichen Effekt. WYSIATI. Der gegenwärtige Gemütszustand spielt eine sehr große Rolle, wenn Menschen ihr Glücksgefühl einschätzen.[5]

## Die Affektheuristik

Die Dominanz von Schlussfolgerungen gegenüber Argumenten ist besonders ausgeprägt, wenn Emotionen im Spiel sind. Der Psychologe Paul Slovic hat eine Affektheuristik vorgeschlagen, wobei Vorlieben und Abneigungen die Weltanschauungen von Menschen bestimmen sollen. Unsere politischen Präferenzen bestimmen, welche Argumente wir überzeugend finden. Wenn man die gegenwärtige Gesundheitspolitik befürwortet, glaubt man, dass sie einen großen Nutzen hat und dass ihre Kosten besser beherrschbar sind als die Kosten von Alternativen. Wenn man in seiner Haltung gegenüber anderen Nationen einen harten außenpolitischen Kurs befürwortet, glaubt man vermutlich, dass sie relativ schwach sind und sich wahrscheinlich dem Willen des eigenen Landes beugen werden. Wenn man einen auf Ausgleich bedachten Kurs befürwortet, glaubt man vermutlich, dass sie stark sind und sich nicht mühelos in die Schranken weisen lassen. Ihre emotionale Einstellung zu Dingen wie bestrahltem Essen, rotem Fleisch, Kernkraft, Tätowierungen oder Motorrädern bestimmt Ihre Ansichten zu deren Nutzen und Risiken. Wenn Sie eine Abneigung gegen eines dieser Dinge haben, glauben Sie voraussichtlich, dass seine Risiken hoch sind und sein Nutzen marginal ist.

Der Vorrang von Schlussfolgerungen bedeutet nicht, dass unser Verstand vollkommen geschlossen ist und dass unsere Meinungen völlig immun gegen Informationen und vernünftiges Nachdenken sind. Unsere Überzeugungen und auch unsere emotionale Einstellung mögen sich ändern (zumindest ein wenig), wenn wir erfahren, dass das Risiko einer Aktivität, die wir nicht mögen, kleiner ist, als wir dachten. Doch die Information über die niedrigeren Risiken wird auch Ihre Sicht des Nutzens (zum Positiven) verändern, selbst wenn in der Information, die Sie erhalten haben, nichts über den Nutzen gesagt wurde.[6]

Wir sehen hier eine neue Seite der »Persönlichkeit« von System 2. Bis jetzt habe ich es überwiegend als einen mehr oder weniger gefügigen Beobachter beschrieben, der System 1 erheblichen Spielraum lässt. Ich habe System 2 auch als aktiv bei der gezielten Gedächtnissuche, komplexen Berechnungen, Vergleichen, Planungen und Entscheidungen beschrieben. Bei dem Schläger-und-Ball-Problem und bei vielen anderen Beispielen, bei

denen die beiden Systeme miteinander wechselwirken, schien es, als hätte System 2 die oberste Zuständigkeit und könnte den Vorschlägen von System 1 widerstehen, kognitive Prozesse verlangsamen und eine logische Analyse auferlegen. Selbstkritik ist eine der Funktionen von System 2. Doch im Kontext von Einstellungen ist System 2 eher ein Fürsprecher der Emotionen von System 1 als ein Kritiker dieser Emotionen – eher ein Unterstützer als ein Kontrolleur. Seine Suche nach Informationen und Argumenten beschränkt sich überwiegend auf Informationen, die mit bestehenden Überzeugungen in Einklang stehen, und verfolgt nicht die Absicht, diese zu überprüfen. Ein aktives, nach Kohärenz strebendes System 1 schlägt einem anspruchslosen System 2 Lösungen vor.

## Zum Thema »Ersetzung und Heuristiken«

»Erinnern wir uns noch an die Frage, die wir beantworten wollen? Oder haben wir sie durch eine leichtere ersetzt?«

»Die eigentliche Frage lautet, ob diese Kandidatin Erfolg haben kann. Die Frage, die wir zu beantworten scheinen, lautet, ob sie bei Interviews gut rüberkommt. Wir sollten die Frage nicht ersetzen.«

»Da er das Projekt mag, glaubt er, dass es niedrige Kosten und einen hohen Nutzen hat. Ein hübsches Beispiel für die Affektheuristik.«

»Wir benutzen die Ertragszahlen des letzten Jahres als eine Heuristik zur Vorhersage des Unternehmenswertes in ein paar

Jahren. Ist diese Heuristik gut genug? Welche weiteren Informationen benötigen wir?«

Die folgende Aufstellung enthält eine Liste von Merkmalen und Aktivitäten, die System 1 zugeschrieben werden. Jeder dieser aktiven Sätze ersetzt eine Aussage, die technisch präziser, aber schwerer zu verstehen wäre, mit der Folge, dass ein mentales Ereignis automatisch und schnell geschieht. Ich hoffe, dass Ihnen die Liste der Merkmale helfen wird, ein intuitives Gefühl für die »Persönlichkeit« des fiktiven Systems 1 zu entwickeln. Wie bei anderen Personen, die Sie kennen, werden Sie intuitiv erahnen, was System 1 unter verschiedenen Umständen tun wird, und die meisten Ihrer Ahnungen werden zutreffend sein.

**Merkmale von System 1**

– erzeugt Eindrücke, Gefühle und Neigungen; unterstützt durch System 2, werden diese zu Überzeugungen, Einstellungen und Intentionen
– arbeitet automatisch und schnell, weitgehend mühelos und ohne willentliche Steuerung
– kann von System 2 so programmiert werden, dass Aufmerksamkeit mobilisiert wird, wenn ein bestimmtes Muster erkannt wird (Suche)
– führt sachkundige Reaktionen aus und erzeugt sachkundige Intuitionen, nach ausreichender Schulung
– erzeugt ein kohärentes Muster aktivierter Vorstellungen im assoziativen Gedächtnis
– verbindet ein Gefühl kognitiver Leichtigkeit mit Wahrheitsillusionen, angenehmen Gefühlen und verminderter Vigilanz
– unterscheidet das Überraschende vom Normalen
– schließt auf Ursachen und Absichten und erfindet beide
– vernachlässigt Ambiguität und unterdrückt Zweifel
– hat die Tendenz, Informationen zu glauben und zu bestätigen
– überzeichnet die emotionale Konsistenz (Halo-Effekt)
– konzentriert sich auf verfügbare Informationen und blendet fehlende aus ( WYSIATI)
– erzeugt eine beschränkte Menge elementarer Bewertungen
– repräsentiert Mengen durch Normen und Prototypen, integriert nicht

– stimmt Intensitäten skalenübergreifend ab (z. B. Größe mit Lautstärke)
– berechnet mehr als beabsichtigt (mentale Schrotflinte)
– ersetzt manchmal eine schwierige Frage durch eine leichtere (Heuristik)
– reagiert empfindlicher auf Veränderungen als auf Zustände (Neue Erwartungstheorie)[a]
– übergewichtet niedrige Wahrscheinlichkeiten[a]
– zeigt verminderte Empfindlichkeit für Quantität (Psychophysik)[a]
– reagiert stärker auf Verluste als auf Gewinne ( Verlustaversion)[a]
– fasst Entscheidungsprobleme eng und getrennt voneinander[a]

# TEIL II

## Heuristiken und kognitive Verzerrungen

# *10. Das Gesetz der kleinen Zahlen*

Eine Studie über die Häufigkeit von Nierenkrebs in den 3141 Countys der Vereinigten Staaten brachte ein bemerkenswertes Muster zum Vorschein. Die Landkreise mit der niedrigsten Nierenkrebshäufigkeit liegen überwiegend in ländlichen, dünn besiedelten und traditionell republikanischen Bundesstaaten im Mittleren Westen, Süden und Westen. Was hat das zu bedeuten?

Ihr Gehirn war in den letzten Sekunden sehr aktiv, und dabei wurde vor allem System 2 beansprucht. Sie haben gezielt Ihr Gedächtnis durchsucht und Hypothesen formuliert. Das war mit einer gewissen Anstrengung verbunden; Ihre Pupillen erweiterten sich, und Ihr Herz schlug messbar schneller. Aber System 1 war nicht untätig: Die Operation von System 2 hing von den Tatsachen und Vorschlägen ab, die aus dem assoziativen Gedächtnis abgerufen wurden. Sie haben vermutlich die Hypothese verworfen, dass die Politik der Republikaner einen Schutz vor Nierenkrebs bietet. Sehr wahrscheinlich konzentrierten Sie sich schließlich auf die Tatsache, dass die Countys mit niedriger Krebshäufigkeit überwiegend ein ländliches Gepräge haben. Die geistreichen Statistiker Howard Wainer und Harris Zwerling, von denen ich dieses Beispiel habe, merkten an: »Es ist sowohl naheliegend als auch verlockend, zu folgern, dass die niedrigeren Krebserkrankungsraten direkt mit der geringen Schadstoffbelastung ländlicher Lebensräume zusammenhängen – keine Luftverschmutzung, keine Wasserverschmutzung, Zugang zu unbelasteten frischen Lebensmitteln.«[1] Dies erscheint vollkommen plausibel.

Betrachten wir jetzt die Landkreise, in denen die Nierenkrebsinzidenz am höchsten ist. Diese »kränkelnden« Countys liegen überwiegend in ländlichen, dünn besiedelten Regionen traditionell republikanischer Bundesstaaten im Mittleren Westen, dem Süden und dem Westen der USA. Augenzwinkernd kommentierten Wainer und Zwerling: »Man könnte leicht folgern, dass die höheren Krebserkrankungsraten direkt auf die Defizite des ländlichen Lebensstils zurückzuführen sind – kein Zugang zu guter medizinischer Versorgung, fettreiche Ernährung und zu viel Alkohol und Tabak.« Selbstverständlich kann da etwas nicht stimmen. Der ländliche

Lebensstil kann nicht zugleich die sehr hohe und die sehr niedrige Häufigkeit von Nierenkrebs erklären.

Der Schlüsselfaktor ist nicht, dass die Landkreise in ländlich geprägten oder überwiegend republikanischen Regionen lagen. Entscheidend ist vielmehr, dass ländliche Countys eine geringe Bevölkerungszahl haben. Und die wichtigste Lektion, die wir daraus lernen können, betrifft nicht die Epidemiologie, sondern die schwierigen Beziehungen zwischen unserem Intellekt und der Statistik. System 1 versteht sich hervorragend auf eine bestimmte Form des Denkens – es erkennt automatisch und mühelos kausale Verknüpfungen zwischen Ereignissen, manchmal sogar dann, wenn diese Beziehung zweifelhaft ist. Als Sie von den Countys mit hohen Krebsraten hörten, nahmen Sie sofort an, dass sich diese Countys aus irgendeinem Grund von anderen Countys unterscheiden, dass es ein Grund geben muss, der diesen Unterschied erklärt. Doch wie wir sehen werden, ist System 1 recht unbeholfen, wenn es mit »bloßen statistischen« Fakten konfrontiert ist, die die Wahrscheinlichkeit von Ergebnissen verändern, diese aber nicht verursachen.

Ein Zufallsereignis entzieht sich, definitionsgemäß, einer Erklärung, aber Mengen von Zufallsereignissen verhalten sich in einer höchst regelhaften Weise. Stellen Sie sich eine mit Murmeln gefüllte Urne vor. Die Hälfte der Murmeln ist rot, die andere Hälfte ist weiß. Stellen Sie sich als Nächstes eine sehr geduldige Person (oder einen Roboter) vor, die blind vier Murmeln aus der Urne zieht, die Anzahl der roten Kugeln in der Stichprobe notiert, die Kugeln wieder zurück in die Urne wirft, und das gleiche Prozedere dann viele Male wiederholt. Wenn man die Ergebnisse addiert, stellt man fest, dass das Ergebnis »2 rot, 2 weiß« (fast exakt) sechsmal so häufig vorkommt wie das Ergebnis »4 rot« oder »4 weiß«. Diese Beziehung ist eine mathematische Tatsache. Wir können das Ergebnis wiederholter Stichprobennahmen aus einer Urne genauso zuverlässig vorhersagen, wie wir prognostizieren können, was geschieht, wenn wir mit einem Hammer auf ein Ei schlagen. Wir können nicht im Einzelnen voraussagen, wie die Schale zerspringen wird, aber wir können sicher sein, *dass* sie zerspringen wird. Es gibt allerdings einen Unterschied: Das Gefühl der Befriedigung, das man erlebt, wenn man einen eindeutigen Kausalzusammenhang erkennt – bei der Vorstellung, dass ein Hammer ein Ei trifft –, fehlt, wenn man an die Ergebnisse statistischer Probennahmen denkt.

Eine damit zusammenhängende statistische Tatsache ist für das Krebsbeispiel von Belang. Zwei sehr geduldige Murmelzähler ziehen abwechselnd aus derselben Urne. Jack zieht bei jeder Probennahme vier Murmeln, Jill zieht sieben Murmeln. Beide notieren sich jedes Mal, wenn eine homogene Stichprobe vorliegt – nur weiß oder nur rot. Wenn sie lange genug weitermachen, wird Jack solche extremen Ergebnisse häufiger beobachten als Jill – und zwar achtmal so häufig (die erwarteten Prozentsätze sind 12,5 und 1,56 Prozent). Wieder kein Hammer, keine kausale Verknüpfung, aber eine mathematische Tatsache: Stichproben aus vier Murmeln führen häufiger zu extremen Ergebnissen als Stichproben mit sieben Murmeln.

Stellen wir uns jetzt die Bevölkerung der Vereinigten Staaten als Murmeln in einer riesigen Urne vor. Einige Murmeln tragen die Markierung NK, für Nierenkrebs. Man zieht Murmel-Stichproben und bevölkert dann nacheinander jedes County. Ländliche Stichproben sind kleiner als andere. Wie in dem Beispiel von Jack und Jill finden sich extreme Ergebnisse (sehr hohe und/oder sehr niedrige Krebsraten) am ehesten in dünn besiedelten Countys. Mehr hat es mit dieser Geschichte nicht auf sich.

Wir gingen von einer Tatsache aus, die nach einer Ursache ruft: Die Nierenkrebshäufigkeit schwankt erheblich zwischen den Countys, und die Unterschiede sind systematischer Natur. Ich habe eine statistische Erklärung vorgeschlagen: Extreme Ergebnisse (sowohl hohe als auch niedrige) kommen eher in kleinen als in großen Stichproben vor. Dies ist keine kausale Erklärung. Die geringe Bevölkerungszahl eines County verursacht keinen Krebs, aber sie schützt auch nicht davor; sie erlaubt lediglich, dass die Krebsrate viel höher (oder viel niedriger) ist als in der allgemeinen Bevölkerung. Die tiefere Wahrheit ist, dass es nichts zu erklären gibt. Die Krebshäufigkeit liegt in einem bevölkerungsarmen County nicht wirklich über oder unter dem allgemeinen Durchschnitt, es scheint in einem bestimmten Jahr aufgrund eines statistischen Zufalls lediglich so zu sein. Wenn wir die Analyse im Jahr darauf wiederholen, werden wir bei den kleinen Stichproben das gleiche allgemeine Muster extremer Ergebnisse beobachten, aber die Countys mit hoher Krebsrate im letzten Jahr werden nicht unbedingt in diesem Jahr wieder eine hohe Krebsrate aufweisen. Wenn dem so ist, zählen die Unterschiede zwischen dicht besiedelten und ländlichen Countys nicht wirklich als Fakten – sie sind vielmehr das, was Wissenschaftler Artefakte nennen, Beobachtungen, die ausschließlich

aufgrund eines bestimmten Aspekts der Forschungsmethode zustande kommen. Im vorliegenden Fall sind dies die unterschiedlichen Stichprobengrößen.

Die Geschichte, die ich erzählt habe, mag Sie überrascht haben, aber es war gewiss keine Offenbarung. Sie wussten bereits, dass die Ergebnisse großer Stichproben vertrauenswürdiger und aussagekräftiger sind als die von kleineren Stichproben, und selbst Menschen, die statistisch wenig bewandert sind, haben schon von diesem Gesetz der großen Zahlen gehört. Aber Wissen ist keine Alles-oder-nichts-Angelegenheit, und Sie werden vielleicht feststellen, dass die folgenden Aussagen auf Sie zutreffen:

- Das Merkmal »dünn besiedelt« war nicht sofort als relevant zu erkennen, als Sie die epidemiologischen Hintergrundinformationen lasen.
- Die Größe des Unterschieds zwischen Vierer- und Siebener-Stichproben hat Sie zumindest leicht erstaunt.
- Selbst jetzt müssen Sie sich mental noch ein wenig anstrengen, um zu erkennen, dass die folgenden beiden Aussagen exakt das Gleiche bedeuten:
  - Große Stichproben liefern präzisere Ergebnisse als kleine Stichproben.
  - Kleine Stichproben führen häufiger zu extremen Ergebnissen als große Stichproben.

Die erste Aussage hört sich wahr an, aber solange Ihnen die zweite Aussage nicht intuitiv einleuchtet, haben Sie die erste nicht richtig verstanden.

Das Fazit: Ja, Sie wussten, dass die Ergebnisse großer Stichproben präziser sind, aber Sie haben jetzt vielleicht erkannt, dass Ihr Wissen nicht besonders profund war. Da sind Sie nicht allein. Die erste Studie, die Amos und ich zusammen durchführten, zeigte, dass selbst fachkundige Forscher ungenaue Intuitionen haben und Stichprobeneffekte nur vage verstehen.

## Das Gesetz der kleinen Zahlen

Meine Zusammenarbeit mit Amos Anfang der 1970er-Jahre begann mit einer Diskussion der Behauptung, Menschen, die nicht in Statistik unterrichtet wurden, seien dennoch gute »intuitive Statistiker«. Er erzählte meinen Studenten und mir von Forschern an der Universität Michigan, die im Allgemeinen die intuitive statistische Kompetenz optimistisch einschätzten. Ich hatte eine entschiedene Meinung zu dieser Behauptung: Ich hatte vor Kurzem entdeckt, dass ich kein guter intuitiver Statistiker war, und ich glaubte nicht, dass ich schlechter als andere war.

Für einen forschenden Psychologen ist die Stichprobenvarianz keine Kuriosität; sie ist ein Ärgernis und ein kostspieliges Hindernis, das jedes Forschungsprojekt zu einem Glücksspiel macht. Angenommen, Sie wollen die Hypothese bestätigen, dass sechsjährige Mädchen im Durchschnitt einen größeren Wortschatz haben als gleichaltrige Jungen. Die Hypothese trifft für die Grundgesamtheit (Gesamtheit aller potenziellen Untersuchungsobjekte für eine Fragestellung) zu; der durchschnittliche Wortschatz von Mädchen ist tatsächlich größer. Innerhalb der Gruppen der Mädchen und Jungen gibt es jedoch große Unterschiede, und der Faktor Zufall beim Auswahlverfahren könnte dafür sorgen, dass man eine Stichprobe auswählt, in der der Unterschied im Wortschatz nicht eindeutig ist, oder auch eine, in der Jungen sogar besser abschneiden. Für einen Forscher ist ein solches Ergebnis kostspielig, weil er umsonst Zeit und Mühe aufgewendet hat und eine Hypothese, die tatsächlich wahr ist, nicht bestätigt hat. Die Auswahl einer hinreichend großen Stichprobe ist der einzige Weg, um dieses Risiko zu vermindern. Forscher, die eine zu kleine Stichprobe wählen, liefern sich dem Glück einer »zufällig« passenden Stichprobe aus.

Das Fehlerrisiko lässt sich für jeden Stichprobenumfang mit einem recht einfachen Verfahren abschätzen. Doch traditionellerweise führen Psychologen keine Berechnungen durch, um den angemessenen Stichprobenumfang zu ermitteln. Sie verlassen sich auf ihr intuitives Urteil, das normalerweise fehlerhaft ist. Ein Artikel, den ich kurz vor der Diskussion mit Amos gelesen hatte, bestätigte den Fehler, den Forscher machen (übrigens bis heute), anhand einer dramatischen Feststellung. Der

Autor wies darauf hin, dass Psychologen im Allgemeinen Stichproben auswählen, die so klein sind, dass sie sich einem fünfzigprozentigen Risiko aussetzen, ihre eigentlich richtigen Hypothesen nicht zu bestätigen![2] Kein Forscher, der recht bei Verstand ist, würde ein derartiges Risiko in Kauf nehmen. Eine plausible Erklärung war, dass sich in den Entscheidungen der Psychologen über den Stichprobenumfang weitverbreitete intuitive Fehlannahmen über die Schwankungsbreite der Ergebnisse in Abhängigkeit von der Stichprobengröße widerspiegelten.

Der Artikel erschütterte mich, weil er einige Schwierigkeiten erklärte, die ich bei meinen Forschungsprojekten gehabt hatte. Wie die meisten Psychologen hatte ich routinemäßig Stichproben ausgewählt, die zu klein waren, und ich hatte dadurch Ergebnisse erhalten, die keinen Sinn ergaben. Jetzt wusste ich, wieso: Die ungewöhnlichen Ergebnisse waren tatsächlich Artefakte meiner Forschungsmethode. Mein Fehler war besonders peinlich, weil ich Statistik unterrichtete und wusste, wie man den Stichprobenumfang berechnet, der das Fehlerrisiko auf ein akzeptables Niveau reduzierte. Aber ich hatte noch nie einen Stichprobenumfang berechnet. Wie meine Kollegen hatte ich bei der Planung meiner Experimente der Tradition und meiner Intuition vertraut und nie ernsthaft über das Problem nachgedacht. Als Amos mein Seminar besuchte, war ich bereits zu dem Schluss gelangt, dass meine Intuition mangelhaft war, und im Verlauf des Seminars erzielten wir rasch Einvernehmen darüber, dass sich die Optimisten in Michigan irrten.

Amos und ich begannen, der Frage nachzugehen, ob ich der einzige Esel war oder Mitglied einer Mehrheit von Eseln, indem wir testeten, ob mathematisch versierte Forscher ähnliche Fehler begehen würden. Wir entwickelten einen Fragebogen, der realistische Forschungssituationen beschrieb, unter anderem auch Wiederholungen erfolgreicher Experimente. Die Forscher wurden aufgefordert, die Stichprobengrößen festzulegen, die Fehlerrisiken zu bewerten, denen sie durch ihre Entscheidungen ausgesetzt waren, und hypothetische Doktoranden bei der Planung ihrer Studien zu beraten. Amos sammelte die Antworten einer Gruppe qualifizierter Teilnehmer (darunter waren auch die Autoren zweier Statistik-Lehrbücher) bei einer Tagung der Gesellschaft für Mathematische Psychologie ein. Die Ergebnisse waren eindeutig: Ich war nicht der einzige Stümper. Jeden der Fehler, die ich gemacht hatte, hatte auch eine große Mehrheit der von uns Befragten gemacht. Ganz offensichtlich schenkten selbst die Experten dem Stichprobenumfang keine hinlängliche Beachtung.

Amos und ich nannten unseren ersten gemeinsamen Artikel »Glaube an das Gesetz der kleinen Zahlen«.[3] Wir erklärten augenzwinkernd, dass »Intuitionen über die Zufallsauswahl von Stichproben das Gesetz der kleinen Zahlen zu erfüllen scheinen, das besagt, dass das Gesetz der großen Zahlen auch für kleine Zahlen gilt«. Wir plädierten auch nachdrücklich dafür, dass Forscher ihren »statistischen Intuitionen mit angemessenem Misstrauen begegnen und, statt sich auf Eindrücke zu verlassen, möglichst Berechnungen durchführen sollten«.[4]

## Die Tendenz, eher zu glauben als zu zweifeln

> Bei einer telefonischen Befragung von 300 Senioren erklärten 60 Prozent ihre Unterstützung für den Präsidenten.

Wenn Sie die Botschaft dieses Satzes in genau drei Wörtern zusammenfassen sollten, wie würden diese lauten? Höchstwahrscheinlich würden Sie »Senioren unterstützen den Präsidenten« äußern. Diese Wörter geben den Kern der Aussage wieder. Die weggelassenen Einzelheiten der Erhebung – dass sie telefonisch durchgeführt wurde und die Stichprobe 300 Personen umfasste – sind, für sich genommen, uninteressant; sie liefern Hintergrundinformationen, die wenig Aufmerksamkeit auf sich ziehen. Ihr Resümee wäre das Gleiche, wenn der Stichprobenumfang ein anderer gewesen wäre. Natürlich würden Sie bei einer völlig absurden Zahl aufmerken (»… eine telefonische Befragung von 6 [oder 60 Millionen] älteren Wählern …«). Aber wenn Sie kein Experte sind, werden Sie auf eine Stichprobe von 150 einerseits und eine von 3000 andererseits nicht sehr unterschiedlich reagieren. Dies ist die Bedeutung der Aussage: »Menschen schenken dem Stichprobenumfang keine hinreichende Beachtung.«

Die Nachricht über die Erhebung enthält zwei verschiedene Typen von Informationen: die Sachinformation an sich und die Informationsquelle. Sie interessieren sich verständlicherweise eher für die Sachinformation als für die Zuverlässigkeit der Ergebnisse. Aber wenn die Zuverlässigkeit offenkundig niedrig ist, verliert die Nachricht ihre Glaubwürdigkeit. Wenn man Ihnen sagt: »Eine Gruppe von Sympathisanten hat eine fehlerhafte und tendenziöse Erhebung durchgeführt, um den Nachweis zu erbringen, dass die Senioren den Präsidenten unterstützen …«, werden Sie natürlich die Ergebnisse der Befragung ablehnen, und sie werden nicht Teil Ihrer Überzeugungen. Stattdessen werden die Sympathisantenumfrage und ihre falschen Ergebnisse eine neue Geschichte über politische Lügen darstellen. In solchen klaren Fällen können Sie frei entscheiden, ob Sie eine Botschaft nicht glauben. Aber unterscheiden Sie auch hinlänglich zwischen »Ich las in

der *New York Times* …« und »Ich hörte beim Kaffeeautomaten …«? Kann Ihr System 1 zwischen Stufen des Überzeugtseins unterscheiden? Das WYSIATI-Prinzip spricht dafür, dass es dies nicht kann.

Wie früher beschrieben, ist System 1 nicht anfällig für Zweifel. Es unterdrückt Ambiguität und konstruiert spontan Geschichten, die so kohärent wie möglich sind. Sofern die Nachricht nicht sofort verneint wird, breiten sich die Assoziationen, die sie hervorruft, so aus, als wäre die Nachricht wahr. System 2 kann zweifeln, weil es gleichzeitig inkompatible Möglichkeiten aufrechterhalten kann. Aber es ist anstrengender, Zweifel aufrechtzuerhalten, als in Gewissheit zu verfallen. Das Gesetz der kleinen Zahlen ist eine Ausprägung der allgemeinen Tendenz, eher zu glauben, als zu zweifeln, die uns in den nächsten Kapiteln in zahlreichen Erscheinungsformen begegnen wird.

Die starke Tendenz, zu glauben, dass kleine Stichproben weitgehend mit der Population übereinstimmen, der sie entnommen wurden, ist ebenfalls Teil einer allgemeineren kognitiven Verzerrung: Wir neigen dazu, die Konsistenz und Kohärenz dessen, was wir sehen, zu überzeichnen. Dieser überzogene Glaube von Forschern an die Aussagekraft von einigen wenigen Beobachtungen ist eng mit dem Halo-Effekt verbunden, dem Gefühl, das wir oftmals haben, einen Menschen, über den wir tatsächlich nur wenig wissen, gut zu kennen und zu verstehen. System 1 greift den Tatsachen vor, wenn es auf der Grundlage von Informationsbruchstücken ein differenziertes Bild konstruiert. Eine Maschine für voreilige Schlussfolgerungen arbeitet so, als würde sie an das Gesetz der kleinen Zahlen glauben. Ganz allgemein bringt sie eine Repräsentation der Wirklichkeit hervor, die mehr Sinn ergibt, als vorhanden ist.

## Ursache und Zufall

Die Assoziationsmaschine sucht nach Ursachen. Die Schwierigkeit, die wir mit statistischen Regelmäßigkeiten haben, besteht darin, dass sie eine andere Herangehensweise erfordern. Statt sich auf die Frage zu konzentrieren, wodurch ein bestimmtes Ereignis hervorgebracht wurde, fragt die statistische Betrachtungsweise, was sich stattdessen hätte ereignen können. Dieses konkrete Ereignis hatte keine bestimmte Ursache – der Zufall hat es unter seinen Alternativen ausgewählt.

Unsere Neigung zu kausalem Denken macht uns anfällig für gravierende Fehler bei der Beurteilung der Zufälligkeit echter Zufallsereignisse. Nehmen wir zum Beispiel das Geschlecht von sechs Säuglingen, die nacheinander in einem Krankenhaus geboren werden. Die Folge von Jungen und Mädchen ist offensichtlich zufallsabhängig; die Ereignisse sind unabhängig voneinander, und die Anzahl der Jungen und Mädchen, die in den letzten Stunden im Krankenhaus zur Welt kamen, hat keinerlei Auswirkungen auf die Geschlechtszugehörigkeit des nächsten Neugeborenen. Betrachten wir jetzt drei mögliche Folgen:

J J J M M M
M M M M M M
J M J J M J

Sind die Folgen gleich wahrscheinlich? Die intuitive Antwort – »Natürlich nicht!« – ist falsch. Weil die Ereignisse unabhängig voneinander sind und weil die Ergebnisse J und M (annähernd) gleich wahrscheinlich sind, ist jede mögliche Folge von sechs Geburten so wahrscheinlich wie jede andere. Selbst jetzt, wo Sie wissen, dass diese Schlussfolgerung wahr ist, bleibt sie kontraintuitiv, weil nur die dritte Folge einen Zufallscharakter aufzuweisen scheint. Erwartungsgemäß wird JMJJMJ als viel wahrscheinlicher beurteilt als die beiden anderen Folgen. Wir suchen unwillkürlich überall nach Mustern, wir glauben fest an eine kohärente Welt, in der Regelmäßigkeiten (wie etwa die Folge von sechs Mädchen) nicht zufällig auftreten, sondern

die Folge mechanischer Kausalität oder der Intention eines Handelnden sind. Wir erwarten nicht, dass Regelmäßigkeit durch einen Zufallsprozess hervorgebracht wird, und wenn wir eine scheinbare Regel erkennen, verwerfen wir schnell die Annahme, der Prozess sei in Wahrheit ein zufallsabhängiger. Zufallsprozesse erzeugen viele Folgen, die Menschen davon überzeugen, dass der Prozess keineswegs zufällig ist. Sie ahnen vermutlich, weshalb die Kausalitätsvermutung evolutionäre Vorteile mit sich bringen kann. Sie ist Bestandteil der allgemeinen Vigilanz – Aufmerksamkeit –, die wir von unseren Ahnen geerbt haben. Wir halten unwillkürlich Ausschau danach, ob sich etwas in der Umwelt verändert hat. Löwen mögen zu unvorhersehbaren Zeiten auf einer Ebene auftauchen, aber es wäre sicherer, wenn man eine scheinbare Zunahme in der Häufigkeit des Auftauchens von Löwenrudeln registrieren und darauf reagieren könnte, auch wenn es sich tatsächlich um Schwankungen bei einem Zufallsprozess handelt.

Das weitverbreitete Missverstehen des Zufallsbegriffs hat manchmal weitreichende Folgen. In unserem Artikel über Repräsentativität zitierten Amos und ich den Statistiker William Feller, der anhand eines Beispiels verdeutlichte, wie leicht wir Muster erkennen, wo gar keine vorhanden sind. Während des intensiven Raketenbeschusses von London im Zweiten Weltkrieg wurde allgemein angenommen, dass der Beschuss nicht wahllos erfolgt sei, weil eine Karte der Einschläge auffällige Lücken zum Vorschein brachte. Einige vermuteten, in den verschonten Gebieten hielten sich deutsche Spione auf.[5] Eine sorgfältige statistische Analyse enthüllte, dass die Verteilung der Einschläge typisch für einen Zufallsprozess war – und auch typisch insofern, als sie den starken Eindruck hervorrief, nicht zufällig zu sein. »Dem ungeübten Auge«, bemerkt Feller, »erscheint eine zufällige Anordnung als Regelmäßigkeit oder Tendenz zur Clusterbildung.«

Ich hatte bald Gelegenheit, das, was ich von Feller gelernt hatte, anzuwenden. Der Jom-Kippur-Krieg brach 1973 aus, und mein einziger substanzieller Beitrag zu den Kriegsanstrengungen war der Ratschlag an hohe Offiziere der israelischen Luftwaffe, eine laufende Untersuchung einzustellen. Der Luftkrieg verlief zunächst wegen der überraschend guten Treffgenauigkeit ägyptischer Boden-Luft-Raketen ziemlich schlecht für Israel. Die Verluste waren hoch, und sie schienen ungleichmäßig verteilt zu sein. Man erzählte mir von zwei Geschwadern, die auf demselben Luftwaffenstützpunkt stationiert waren und von denen das eine vier

Flugzeuge verloren hatte, während das andere keine Verluste zu beklagen hatte. Es wurde eine Untersuchung eingeleitet, um in Erfahrung zu bringen, was das vom Pech verfolgte Geschwader falsch machte. Es gab im Vorfeld keine Anhaltspunkte dafür, dass eines der beiden Geschwader leistungsfähiger war als das andere, und es wurden keine Unterschiede bei der Einsatzfähigkeit festgestellt. Aber natürlich unterschieden sich die Lebensweisen der Piloten in vielfältiger zufallsabhängiger Weise, unter anderem auch darin, wie ich mich erinnere, wie oft sie zwischen den Einsätzen nach Hause fuhren und wie die Nachbesprechungen geführt wurden. Mein Rat an das Oberkommando lautete, sie sollten sich damit abfinden, dass die verschiedenen Ergebnisse rein zufällig zustande gekommen seien, und sie sollten die Befragung der Piloten einstellen. Ich argumentierte, dass der Zufall die wahrscheinlichste Antwort sei, dass eine Suche ins Blaue hinein nach einer verborgenen Ursache aussichtslos sei und dass in der Zwischenzeit die Piloten in dem Geschwader, das Verluste erlitten hatte, nicht der zusätzlichen Belastung ausgesetzt werden sollten, ihnen das Gefühl zu geben, sie und ihre toten Freunde hätten sich falsch verhalten.

Einige Jahre später erregten Amos und seine Studenten Tom Gilovich und Robert Vallone mit ihrer Studie über Fehleinschätzungen von Zufallsereignissen im Basketball großes Aufsehen.[6] Die »Tatsache«, dass Spieler hin und wieder eine glückliche Hand haben, wird von Spielern, Trainern und Fans allgemein anerkannt. Die Schlussfolgerung daraus scheint unwiderstehlich zu sein: Ein Spieler schafft es, drei oder vier Bälle hintereinander im Korb zu versenken, sodass sich einem das spontane Urteil aufdrängt, dass dieser Spieler jetzt eine glückliche Hand hat, mit einer vorübergehend erhöhten Treffsicherheit. Spieler beider Mannschaften passen ihr Verhalten an dieses Urteil an – Teamkollegen spielen den Ball eher dem Torjäger zu, und die Abwehr wird ihn eher decken. Doch Analysen von Tausenden von Torwürfen erbrachten ein ernüchterndes Ergebnis: Im Profibasketball gibt es nichts dergleichen wie eine glückliche Hand, weder beim Werfen vom Spielfeld noch beim Werfen von der Freiwurflinie. Natürlich sind einige Spieler treffsicherer als andere, aber die Folge von Treffern und Fehlwürfen besteht jeden Zufallstest. Die glückliche Hand liegt zur Gänze im Auge der Betrachter, die immer allzu schnell in zufälligen Ereignissen Ordnung und Kausalität erkennen. Die glückliche Hand ist eine mächtige und weitverbreitete kognitive Illusion.

Die öffentliche Reaktion auf diese Studie ist Teil der Geschichte. Die Ergebnisse wurden wegen ihres überraschenden Charakters von der Presse aufgegriffen, und die allgemeine Reaktion war Ungläubigkeit. Als der gefeierte Trainer der Boston Celtics, Red Auerbach, von Gilovich und seiner Studie hörte, antwortete er: »Wer ist dieser Kerl? Hat also eine Studie gemacht. Das ist mir völlig schnuppe.« Die Tendenz, Muster in Zufallsereignissen zu erkennen, ist einfach unwiderstehlich – sie ist gewiss eindrucksvoller als irgend so ein Typ, der eine Studie macht.

Die Musterillusion wirkt sich auch außerhalb des Basketballfeldes in vielfältiger Weise auf unser Leben aus. Wie viele gute Jahre sollte man warten, bis man den Schluss zieht, dass ein Vermögensberater ungewöhnlich tüchtig ist? Wie viele erfolgreiche Zukäufe sind nötig, damit ein Verwaltungsrat den sachlich begründeten Schluss ziehen kann, dass der Vorstandschef eine außergewöhnliche Begabung für solche Übernahmen hat? Die einfache Antwort auf diese Frage lautet, dass man, wenn man sich auf seine Intuition verlässt, häufig irrt, indem man ein Zufallsereignis als ein systematisches fehlinterpretiert. Wir lehnen nur allzu bereitwillig die Ansicht ab, dass ein Großteil dessen, was wir im Leben sehen, Zufallscharakter hat.

Ich begann dieses Kapitel mit dem Beispiel der Krebshäufigkeit in den Vereinigten Staaten. Dieses Beispiel stammt aus einem Buch für Statistiklehrer, aber ich erfuhr davon in einem amüsanten Artikel der beiden Statistiker, die ich weiter vorn zitiert habe, Howard Wainer und Harris Zwerling. Ihr Aufsatz konzentrierte sich auf eine große Summe, etwa 1,7 Milliarden Dollar, die die Gates-Stiftung ausgab, um verblüffenden Ergebnissen über die Merkmale der erfolgreichsten Schulen Rechnung zu tragen. Viele Forscher haben versucht, dadurch hinter das Geheimnis erfolgreicher Bildung zu kommen, dass sie die erfolgreichsten Schulen identifizierten, in der Hoffnung, herauszufinden, was diese von anderen unterscheidet. Eines der Ergebnisse dieser Forschungen ist, dass die erfolgreichsten Schulen, im Schnitt, klein sind. In einer Studie an 1662 Schulen in Pennsylvania zum Beispiel waren sechs der fünfzig besten Schulen klein, sodass sie vierfach überrepräsentiert waren. Diese Daten veranlassten die Gates-Stiftung dazu, erhebliche Finanzmittel für die Schaffung kleiner Schulen bereitzustellen, gelegentlich auch durch Aufspaltung großer Schulen in kleinere Einheiten. Mindestens ein halbes Dutzend andere bekannte Institutionen, wie die Annenberg Foundation und

der Pew Charitable Trust, schlossen sich dem Projekt an, ebenso wie das Smaller Learning Communities Program des US-Bildungsministeriums.

Dies erscheint Ihnen vermutlich intuitiv als sinnvoll. Man kann leicht eine kausale Geschichte konstruieren, die erklärt, wieso kleine Schulen bessere Bildungsergebnisse und daher mehr leistungsstarke Schüler produzieren, nämlich dadurch, dass den Schülern mehr persönliche Aufmerksamkeit und Ermunterung durch die Lehrer zuteilwird, als dies in größeren Schulen möglich wäre. Leider ist diese kausale Analyse müßig, weil die Fakten falsch sind. Wenn die Statistiker, die im Auftrag der Gates-Stiftung tätig waren, nach den Merkmalen der schlechtesten Schulen gefragt hätten, hätten sie festgestellt, dass die schlechten Schulen ebenfalls unterdurchschnittlich klein sind. Die Wahrheit ist, dass kleine Schulen im Durchschnitt nicht besser sind; sie sind lediglich variabler. Wenn überhaupt, so Wainer und Zwerling, dann produzieren große Schulen bessere Leistungen, insbesondere in den höheren Klassen, wo ein reichhaltiges Fächerangebot zur Auswahl steht.

Dank jüngster Fortschritte in der kognitiven Psychologie können wir heute klar erkennen, was Amos und ich nur erahnten: Das Gesetz der kleinen Zahlen ist Teil zweier umfassenderer Erkenntnisse über die Funktionsweise des menschlichen Geistes:

- Der überzogene Glaube an die Aussagekraft kleiner Stichprobenbefunde ist nur ein Beispiel für eine allgemeinere Illusion – wir schenken dem Inhalt von Nachrichten mehr Aufmerksamkeit als der Information über ihre Zuverlässigkeit und gelangen so notwendigerweise zu einer Sicht der Welt um uns herum, die einfacher und kohärenter ist, als es die Daten rechtfertigen. Voreilige Schlussfolgerungen sind in der Welt der Imagination ein sichererer Sport als in der Wirklichkeit.
- Die Statistik bringt viele Beobachtungen hervor, die eine kausale Erklärung nahezulegen scheinen, aber für eine solche Erklärung nicht geeignet sind. Viele Tatsachen der Welt sind auf Zufälle zurückzuführen, einschließlich Zufällen bei der Stichprobenauswahl. Kausale Erklärungen von Zufallsereignissen sind zwangsläufig falsch.

## Zum Thema »Gesetz der kleinen Zahlen«

»Ja, das Filmstudio hatte drei erfolgreiche Filme, seit der neue Geschäftsführer sein Amt angetreten hat. Aber es ist zu früh, um zu erklären, dass er eine glückliche Hand hat.«

»Ich werde nicht glauben, dass der neue Wertpapierhändler ein Genie ist, bevor ich einen Statistiker gefragt habe, der die Wahrscheinlichkeit abschätzen könnte, dass seine Glückssträhne ein Zufallsereignis ist.«

»Die Stichprobe beobachteter Werte ist zu klein, um irgendwelche Schlüsse daraus zu ziehen. Folgen wir nicht dem Gesetz der kleinen Zahlen.«

»Ich will die Ergebnisse des Experiments so lange geheim halten, bis wir eine hinreichend große Stichprobe haben. Andernfalls wird man uns unter Druck setzen, vorschnell zu einem Ergebnis zu kommen.«

# *11. Anker*

Amos und ich haben einmal ein Glücksrad manipuliert. Es war mit einer Markierung von 0 bis 100 versehen, aber wir hatten es so konstruiert, dass es nur auf 10 oder auf 65 stehen blieb. Wir rekrutierten Studenten der Universität Oregon als Teilnehmer unseres Experiments. Einer von uns stand vor einer kleinen Gruppe, drehte das Rad und forderte sie auf, die Zahl aufzuschreiben, bei der das Rad stehen blieb, was natürlich entweder bei der 10 oder der 65 der Fall war. Dann stellten wir ihnen zwei Fragen:

> Ist der Prozentsatz afrikanischer Staaten unter den Mitgliedstaaten der Vereinten Nationen größer oder kleiner als die Zahl, die Sie gerade aufgeschrieben haben?
>
> Wie hoch ist Ihrer Einschätzung nach der Prozentsatz afrikanischer Staaten in den Vereinten Nationen?

Das Drehen eines Glücksrads – selbst eines, das nicht manipuliert wurde – kann keinerlei nützliche Informationen über irgendetwas liefern, und die Teilnehmer unseres Experiments hätten es schlichtweg ignorieren sollen. Aber sie ignorierten es nicht. Die mittleren Schätzwerte derjenigen, die 10 bzw. 65 sahen, beliefen sich auf jeweils 25 bzw. 45 Prozent.

Das Phänomen, das wir studierten, ist im Alltag so weit verbreitet und so wichtig, dass Sie seinen Namen kennen sollten: Es ist ein »Ankereffekt«. Er ereignet sich, wenn Menschen einen bestimmten Wert für eine unbekannte Größe erwägen, bevor sie diese Größe abgeschätzt haben. Was hier geschieht, ist ein Beispiel für einen der zuverlässigsten und robustesten Befunde der experimentellen Psychologie: Die Schätzwerte bleiben nahe bei der Zahl, die den Personen im Vorfeld dargeboten wurde. Wenn man Sie fragt, ob Gandhi über 114 Jahre alt war, als er starb, werden Sie sein Alter bei seinem Tod viel höher schätzen, als Sie es tun würden, wenn die Ankerfrage auf einen Tod mit 35 Jahren verweisen würde. Wenn Sie

überlegen, wie viel Sie für ein Haus bezahlen sollten, werden Sie von der ursprünglichen Preisforderung beeinflusst. Dasselbe Haus wird Ihnen wertvoller erscheinen, wenn sein Listenpreis höher ist, als wenn er niedrig ist, selbst wenn Sie entschlossen sind, sich nicht von dieser Zahl beeinflussen zu lassen; und so weiter – die Liste der Ankereffekte ist endlos. Jede Zahl, die Ihnen als mögliche Lösung für ein Schätzungsproblem präsentiert wird, erzeugt einen Ankereffekt.

Wir waren nicht die Ersten, die die Effekte von Ankern beobachteten, aber unser Experiment wies erstmals nach, wie absurd Ankereffekte sind: Die Urteile von Menschen werden von einer Zahl beeinflusst, die offenkundig keinen Informationsgehalt hat. Der Ankereffekt eines Glücksrads lässt sich in keiner Weise als rational beschreiben. Amos und ich veröffentlichten das Experiment in unserem *Science*-Aufsatz, und es ist einer der bekanntesten Befunde, über die wir dort berichteten.

Es gab nur ein Problem: Amos und ich waren uns über die psychologische Erklärung des Ankereffekts nicht einig. Er befürwortete eine Interpretation, ich zog eine andere vor, und wir konnten unsere Kontroverse nicht beilegen. Das Problem wurde schließlich Jahrzehnte später durch die Anstrengungen zahlreicher Forscher gelöst. Heute steht fest, dass wir beide, Amos und ich, recht hatten. Zwei verschiedene Mechanismen erzeugen Ankereffekte – einer für jedes System. Eine Form der Verankerung geschieht in einem willentlichen Prozess der Anpassung und ist eine Operation von System 2. Eine andere Verankerung vollzieht sich durch einen Priming-Effekt und ist eine automatische Manifestation von System 1.

## Ankerung als Anpassung

Amos gefiel die Idee einer Anpassungs-und-Anker-Heuristik als einer Strategie zur Abschätzung ungewisser Größen: Man geht von einer Ankerzahl aus, schätzt ein, ob sie zu hoch oder zu niedrig ist, und passt seine Schätzung dann nach und nach an, indem man sich mental vom Anker »wegbewegt«. Die Anpassung endet in der Regel verfrüht, weil Menschen innehalten, wenn sie sich nicht mehr sicher sind, ob sie sich weiterbewegen sollten. Jahrzehnte nach unserer Kontroverse und Jahre nach Amos' Tod legten zwei Psychologen, die zu Beginn ihrer wissenschaftlichen Laufbahn eng mit Amos zusammengearbeitet hatten, überzeugende empirische Belege für einen solchen Prozess vor: Eldar Shafir und Tom Gilovich, in Zusammenarbeit mit ihren eigenen Studenten – den intellektuellen Enkeln von Amos!

Um die Idee zu verstehen, nehmen Sie ein Blatt Papier und zeichnen ohne Lineal eine 6,5 Zentimeter lange Linie vom unteren Rand der Seite aufwärts. Nehmen Sie nun ein anderes Blatt, setzen Sie am oberen Rand an und ziehen Sie eine Linie nach unten, bis sie 6,5 Zentimeter vom unteren Rand entfernt ist. Vergleichen Sie die Linien. Die Wahrscheinlichkeit ist hoch, dass Ihre erste Schätzung von 6,5 Zentimetern kürzer ist als die zweite. Der Grund dafür ist, dass Sie nicht genau wissen, wie eine solche Linie aussieht; es gibt eine Unsicherheitsspanne. Wenn Sie unten auf der Seite beginnen, hören Sie in der Nähe des unteren Endes des Unsicherheitsbereichs auf; wenn Sie oben auf der Seite beginnen, ziehen Sie den Strich bis zum oberen Ende dieses Bereichs durch. Robyn LeBoeuf und Shafir fanden im Alltag viele Beispiele für diesen Mechanismus. Ungenügende Anpassung erklärt perfekt, weshalb Sie vermutlich zu schnell fahren, wenn Sie von einer Autobahn auf eine Landstraße kommen – insbesondere, wenn Sie sich beim Fahren mit jemandem unterhalten. Unzureichende Anpassung ist auch eine Ursache für Spannungen zwischen verzweifelten Eltern und Teenagern, die in ihren Zimmern laute Musik hören. LeBoeuf und Shafir bemerken dazu: »Ein wohlmeinendes Kind, das ungewöhnlich laute Musik leiser stellt, um der Forderung der Eltern nachzukommen, Musik bei ›zumutbarer‹ Lautstärke zu hören, korrigiert einen hohen Anker möglicherweise nicht ausreichend und hat vielleicht das

Gefühl, dass echte Kompromissversuche nicht als solche wahrgenommen werden.«[1] Der Fahrer und das Kind korrigieren gezielt nach unten, aber sie korrigieren nicht ausreichend.

Betrachten wir jetzt diese Fragen:

> Wann wurde George Washington Präsident?
> Wie hoch ist die Siedetemperatur von Wasser auf dem Gipfel des Mount Everest?

Wenn Sie über diese Fragen nachdenken, fällt Ihnen als Erstes ein Anker ein, und Sie wissen sowohl, dass er falsch ist, als auch die Richtung der richtigen Antwort. Sie wissen auf Anhieb, dass George Washington erst nach der Unterzeichnung der amerikanischen Unabhängigkeitserklärung 1776 Präsident werden konnte, und Sie wissen auch, dass die Siedetemperatur von Wasser auf dem Mount Everest unter hundert Grad Celsius liegt. Sie müssen in die richtige Richtung korrigieren, indem Sie Argumente finden, um sich vom Anker zu entfernen. Wie bei den Linien werden Sie vermutlich anhalten, wenn Sie nicht sicher sind, ob Sie sich weiterbewegen sollten – am nahen Rand der Unsicherheitsregion.
Nick Epley und Tom Gilovich fanden Belege dafür, dass die Anpassung ein bewusster Versuch ist, Gründe zu finden, um sich vom Anker zu entfernen: Versuchspersonen, die angewiesen werden, ihren Kopf zu schütteln, wenn sie den Anker hören, als würden sie diesen ablehnen, bewegen sich weiter vom Anker weg, und Menschen, die nicken, zeigen eine verstärkte Ankerung.[2] Epley und Gilovich bestätigten auch, dass die Anpassung ein mühsamer Prozess ist. Menschen korrigieren weniger stark (bleiben näher beim Anker), wenn ihre mentalen Ressourcen erschöpft sind, entweder weil ihr Gedächtnis mit Ziffern befrachtet ist oder weil sie leicht betrunken sind.[3] Unzureichende Anpassung ist das Versagen eines schwachen oder trägen Systems 2.

Also wissen wir heute, dass Amos zumindest in einigen Fällen von Ankereffekten recht hatte, die mit einer willentlichen System-2-Anpassung in eine bestimmte Richtung von einem Anker weg verbunden sind.

## Ankerung als ein Priming-Effekt

Als Amos und ich über das Phänomen der Ankerung diskutierten, stimmte ich ihm zu, dass es gelegentlich zu Anpassungen kommt, aber ich fühlte mich unwohl dabei. Die Anpassung ist eine willentliche und bewusste Aktivität, aber in den meisten Fällen von Ankerung gibt es kein entsprechendes subjektives Erleben. Betrachten wir diese beiden Beispiele:

> War Gandhi mehr oder weniger als 144 Jahre alt, als er starb?
> Wie alt war Gandhi, als er starb?

Sind Sie dadurch zu Ihrem Schätzwert gelangt, dass Sie 144 nach unten korrigierten? Vermutlich nicht, und dennoch hat sich die absurd hohe Zahl auf Ihren Schätzwert ausgewirkt. Ich vermutete, dass Ankerung ein Fall von Suggestion ist. Wir benutzen dieses Wort, wenn uns jemand allein dadurch, dass er uns etwas ins Bewusstsein ruft, dazu bringt, dieses Etwas zu sehen, zu hören oder zu spüren. So veranlasst die Frage »Spüren Sie jetzt eine leichte Taubheit in Ihrem linken Bein?« nicht wenige Menschen dazu, zu berichten, dass sich ihr linkes Bein tatsächlich ein wenig seltsam anfühlt.

Amos war gegenüber intuitiven Ahnungen zurückhaltender als ich, und er wies zu Recht darauf hin, dass der Rückgriff auf Suggestion uns nicht dabei helfe, Ankerung zu verstehen, weil wir nicht wüssten, wie wir Suggestion erklären sollten. Ich musste zugeben, dass er recht hatte, aber die Vorstellung, ungenügende Anpassung sei die einzige Ursache von Ankereffekten, hat mir nie so richtig behagt. In dem Bemühen, die Ankerung zu verstehen, führten wir viele Experimente durch, aber die Ergebnisse waren nicht schlüssig, und so gaben wir schließlich den Plan auf, mehr darüber zu schreiben.

Das Problem, an dem wir scheiterten, ist mittlerweile gelöst, weil das Konzept der Suggestion nicht länger unerklärlich ist: Suggestion ist ein Priming-Effekt, der selektiv kompatible Informationen ins Gedächtnis ruft. Sie haben keinen Augenblick lang geglaubt, dass Gandhi 144 Jahre alt wurde, aber Ihre Assoziationsmaschine erzeugte sicherlich einen Eindruck

von einem sehr alten Menschen. System 1 interpretiert Sätze, indem es versucht, sie als wahr hinzustellen, und die selektive Aktivierung kompatibler Gedanken erzeugt eine Reihe systematischer Fehler, die uns leichtgläubig und anfällig dafür machen, unseren Überzeugungen allzu sehr zu vertrauen. Wir können jetzt sehen, wieso Amos und ich nicht erkannten, dass es zwei Typen von Ankereffekten gibt: Die Forschungsmethoden und theoretischen Konzepte, die wir brauchten, waren noch nicht vorhanden. Sie wurden viel später von anderen entwickelt. Tatsächlich läuft in vielen Situationen ein der Suggestion ähnlicher Prozess ab: System 1 tut sein Bestes, um eine Welt zu konstruieren, in welcher der Anker die richtige Zahl ist. Dies ist eine der Manifestationen assoziativer Kohärenz, die ich im ersten Teil des Buches beschrieb.

Die deutschen Psychologen Thomas Mussweiler und Fritz Strack lieferten die überzeugendsten Beweise für die Rolle assoziativer Kohärenz bei der Ankerung. [4] In einem Experiment stellten sie eine Ankerfrage über die Temperatur: »Ist die Jahresdurchschnittstemperatur in Deutschland höher oder niedriger als zwanzig Grad Celsius?« oder »Ist die Jahresdurchschnittstemperatur in Deutschland höher oder niedriger als fünf Grad Celsius?«.

Allen Teilnehmern wurden anschließend kurz Wörter dargeboten, die sie erkennen sollten. Die Forscher fanden heraus, dass »zwanzig Grad Celsius« das Erkennen von Wörtern mit Sommer-Bezug (wie »Sonne« und »Strand«) erleichterte, während »fünf Grad Celsius« das Erkennen von Wörtern, die mit Winter assoziiert sind (wie »Frost« und »Ski«), erleichterte. Die selektive Aktivierung kompatibler Gedächtnisinhalte erklärt die Ankerung: Die hohe und die niedrige Zahl aktivieren verschiedene Vorstellungskomplexe im Gedächtnis. Die Schätzwerte für die Jahrestemperatur leiten sich von diesen verzerrten Vorstellungsstichproben her und sind daher ebenfalls verzerrt. In einer weiteren eleganten Studie ähnlicher Art wurden die Probanden nach dem Durchschnittspreis deutscher Autos gefragt. Ein hoher Anker primte selektiv die Namen von Luxusmarken (Mercedes, Audi), während der niedrige Anker Automarken primte, die mit dem Massenmarkt assoziiert sind (VW). Wir sahen früher, dass jeder Prime tendenziell Informationen evoziert, die mit ihm kompatibel sind. Suggestion und Ankerung werden beide durch den gleichen automatischen Funktionsmechanismus von System 1 erklärt. Obgleich ich damals nicht wusste, wie ich es beweisen sollte, erwies sich

meine Intuition über den Zusammenhang zwischen Ankerung und Suggestion als zutreffend.

## Der Ankerungsindex

Viele psychologische Phänomene lassen sich experimentell demonstrieren, aber nur wenige sind tatsächlich messbar. Der Ankereffekt ist eine Ausnahme. Ankerung ist messbar, und sie ist ein beeindruckend starker Effekt. Einigen Besuchern des San Francisco Exploratorium (ein naturwissenschaftliches Museum) wurden die folgenden beiden Fragen gestellt:[5]

> Beträgt die Höhe des größten Küstenmammutbaums mehr oder weniger als 366 Meter?
> Wie hoch ist Ihrer Meinung nach der größte Küstenmammutbaum?

Der »hohe Anker« in diesem Experiment war 366 Meter. Bei anderen Versuchspersonen bezog sich die erste Frage auf einen »niedrigen Anker« von 55 Metern. Die Differenz zwischen den beiden Ankern betrug 311 Meter.

Erwartungsgemäß wichen die mittleren Schätzwerte der beiden Gruppen stark voneinander ab: 257 und 86 Meter. Die Differenz zwischen beiden betrug 171 Meter. Der Ankerungsindex ist einfach das Verhältnis der beiden Differenzen (171/311), ausgedrückt als Prozentsatz: 55 Prozent. Das Ankerungsmaß betrüge 100 Prozent für Menschen, die sklavisch den Anker als Schätzwert verwendeten, und null bei Personen, die in der Lage sind, den Anker vollkommen zu ignorieren. Der Wert von 55 Prozent, der in diesem Beispiel beobachtet wurde, ist typisch. Ähnliche Werte wurden bei vielen anderen Problemen beobachtet.

Der Ankereffekt ist keine Kuriosität aus Forschungslabors; er kann in der realen Welt genauso wirkmächtig sein. In einem Experiment, das vor einigen Jahren durchgeführt wurde, erhielten Immobilienmakler die Gelegenheit, den Wert eines Hauses einzuschätzen, das tatsächlich zum Verkauf stand. Sie besichtigten das Haus und studierten eine umfangreiche

Informationsbroschüre, die eine Preisforderung enthielt. Die Hälfte der Makler sah eine Preisforderung, die erheblich über dem Listenpreis des Hauses lag; die andere Hälfte sah eine, die deutlich darunterlag.[6] Jeder Makler äußerte seine Meinung über einen angemessenen Kaufpreis für das Haus und den niedrigsten Preis, zu dem er das Haus verkaufen würde, wenn er der Eigentümer wäre. Anschließend wurden die Makler nach den Faktoren gefragt, die sich auf ihr Urteil auswirkten. Bemerkenswerterweise war die Preisforderung keiner dieser Faktoren; die Makler waren stolz auf ihre Fähigkeit, diese zu ignorieren. Sie beteuerten, die Preisforderung habe ihre Antworten nicht beeinflusst, aber sie irrten sich: Der Ankereffekt betrug 41 Prozent. Tatsächlich waren die fachkundigen Praktiker fast genauso anfällig für Ankereffekte wie BWL-Studenten ohne Immobilienerfahrung, deren Ankerungsindex 48 Prozent betrug. Der einzige Unterschied zwischen den beiden Gruppen bestand darin, dass die Studenten einräumten, von dem Anker beeinflusst worden zu sein, während die Profis diesen Einfluss bestritten.

Mächtige Ankereffekte finden sich bei Entscheidungen in Geldfragen, etwa wenn wir überlegen, wie viel wir für eine Sache spenden. Um diesen Effekt nachzuweisen, erzählten wir den Teilnehmern der Exploratorium-Studie in San Francisco von den Umweltschäden, die durch Öltanker im Pazifik verursacht werden, und fragten nach ihrer Bereitschaft, eine jährliche Geldspende zu leisten, »um 50 000 Seevögel an der Pazifikküste vor den Folgen kleinerer Ölteppiche in Küstennähe zu schützen, bis Möglichkeiten gefunden werden, um Ölkatastrophen zu verhindern oder die Eigner von Tankern zur Übernahme der Kosten zu zwingen«. Diese Frage erfordert eine Intensitätsabstimmung: Tatsächlich werden die Probanden aufgefordert, die Höhe eines Spendenbetrags in Dollar zu finden, die der Intensität ihres Mitgefühls mit der Not der Seevögel entspricht. Einigen der Besucher wurde eine Ankerfrage gestellt, wie etwa: »Wären Sie bereit, 5 Dollar zu bezahlen …«, bevor sie rundheraus gefragt wurden, wie viel sie auszugeben bereit wären.

Wenn kein Anker erwähnt wurde, sagten die Besucher des Exploratoriums – im Allgemeinen umweltbewusste Personen –, sie wären bereit, 64 Dollar zu zahlen (der Durchschnittswert). Wenn der Ankerbetrag sich auf nur 5 Dollar belief, lag der durchschnittliche Spendenbetrag bei 20 Dollar. Als der Anker sich auf eher extravagante 400 Dollar belief, stieg die Spendenbereitschaft auf durchschnittlich 143 Dollar.

Die Differenz zwischen den Gruppen mit dem hohen und dem niedrigen Anker betrug 123 Dollar. Der Ankereffekt betrug über 30 Prozent, was darauf hindeutet, dass die Steigerung der anfänglichen Forderung um 100 Dollar eine Rendite von 30 Dollar in der durchschnittlichen Zahlungsbereitschaft erbrachte.

Ähnliche oder auch größere Ankereffekte wurden bei zahlreichen Studien über Schätzwerte und Zahlungsbereitschaft ermittelt. So wurden zum Beispiel Franzosen, die im stark umweltverschmutzten Großraum Marseilles leben, gefragt, welche Erhöhung der Lebenshaltungskosten sie zu akzeptieren bereit wären, wenn sie dafür in einer weniger stark verschmutzten Region leben könnten. In dieser Studie betrug der Ankereffekt über 50 Prozent. Ankereffekte lassen sich auch bei Online-Wertpapiertransaktionen beobachten, wo das gleiche Produkt oftmals zu verschiedenen »Jetzt-kaufen-Preisen« angeboten wird. Die Taxe bei Kunstauktionen ist ebenfalls ein Anker, der das erste Gebot beeinflusst.

Es gibt Situationen, in denen Ankerung vernünftig erscheint. Schließlich ist es nicht überraschend, dass Menschen, denen schwierige Fragen gestellt werden, nach Strohhalmen greifen, und der Anker ist ein plausibler Strohhalm. Wenn man so gut wie nichts über die Bäume Kaliforniens weiß und gefragt wird, ob ein Küstenmammutbaum höher als 366 Meter werden kann, könnte man zu dem Schluss gelangen, dass diese Zahl nicht allzu weit von der Wahrheit entfernt ist. Jemand, der die wahre Länge kennt, hat sich diese Frage ausgedacht, sodass der Anker einen nützlichen Hinweis liefern mag. Doch eine zentrale Erkenntnis der Ankerforschung besteht darin, dass offenkundig willkürlich festgesetzte Anker genauso wirksam sein können wie potenziell informative. Als wir ein Glücksrad benutzten, um Schätzwerte für den Anteil afrikanischer Staaten an den Vereinten Nationen zu verankern, lag der Ankerindex bei 44 Prozent, deutlich innerhalb des Bereichs der Effekte, in dem man Anker beobachtet, die als plausible Hinweise angesehen werden können. Ankereffekte ähnlicher Größe wurden bei Experimenten beobachtet, in denen die letzten Ziffern der Sozialversicherungsnummer von Versuchspersonen als Anker verwendet wurden (zum Beispiel, um die Zahl der Ärzte in ihrer Stadt zu schätzen). Die Schlussfolgerung ist klar: Die Effekte von Ankern sind nicht darauf zurückzuführen, dass Menschen sie für informativ halten.

Einige Experimente haben auf verstörende Weise die Wirkmächtigkeit zufällig gewählter Anker demonstriert. Deutsche Richter mit einer

durchschnittlichen Berufserfahrung von über 15 Jahren lasen die Beschreibung einer Frau, die bei einem Ladendiebstahl erwischt worden war, und warfen dann zwei Würfel, die gezinkt waren, sodass jeder Wurf entweder zu einer Drei oder einer Neun führte.[7] Sobald die Würfel zum Liegen kamen, wurden die Richter gefragt, ob sie die Frau zu einer Freiheitsstrafe verurteilen würden, die, in Monaten, größer oder kleiner als die auf dem Würfel angezeigte Zahl wäre. Schließlich sollten die Richter die Freiheitsstrafe, zu der sie die Ladendiebin verurteilen würden, genau angeben. Diejenigen, die eine Neun gewürfelt hatten, sagten im Schnitt, dass sie die Diebin zu acht Monaten verurteilen würden; diejenigen, die eine Drei gewürfelt hatten, sagten, dass sie sie zu fünf Monaten verurteilen würden; der Ankereffekt betrug 50 Prozent.

## Gebrauch und Missbrauch von Ankern

Mittlerweile sollten Sie überzeugt davon sein, dass Ankereffekte – manchmal aufgrund von Priming, manchmal aufgrund ungenügender Anpassung – allgegenwärtig sind. Die psychologischen Mechanismen, die Ankereffekte erzeugen, machen uns viel empfänglicher für Suggestionen, als es den meisten von uns lieb ist. Und natürlich gibt es eine Reihe von Leuten, die gewillt und in der Lage sind, unsere Leichtgläubigkeit auszunutzen.

Ankereffekte erklären zum Beispiel, weshalb willkürliche Rationierung eine erfolgreiche Marketingstrategie ist. Vor ein paar Jahren konnten Kunden eines Supermarkts in Sioux City, Indiana, Suppen von Campbell's im Rahmen einer Verkaufsförderungsaktion für etwa 10 Prozent des regulären Preises kaufen. An manchen Tagen wurde am Regal ein Schild mit folgender Aufschrift angebracht: »Maximal zwölf Dosen pro Person«. An anderen Tagen wurde ein Schild mit dieser Aufschrift angebracht: »Keine Begrenzung pro Person«.[8] Mit der Beschränkung kauften die Kunden durchschnittlich sieben Dosen – doppelt so viele, wie sie kauften, als es keine Beschränkung gab. Ankerung ist nicht die einzige Erklärung. Rationierung bedeutet auch, dass sich die Regale im Eiltempo leeren, sodass die Käufer eine gewisse Dringlichkeit spüren sollten, sich einzudecken. Aber wir wissen auch, dass die bloße Erwähnung von zwölf Dosen als eine mögliche Einkaufsmenge auch dann einen Ankereffekt erzeugen würde, wenn die Zahl durch ein Rouletterad generiert worden wäre.

Die gleiche Strategie sehen wir bei Verhandlungen über den Preis eines Hauses, wenn der Verkäufer den ersten Zug macht, indem er den Listenpreis nennt. Wie bei vielen anderen Spielen ist es auch bei Verhandlungen, bei denen man sich nur über einen Punkt einigen muss – zum Beispiel über den Preis –, ein Vorteil, wenn man die Initiative ergreift. Wie Sie vielleicht schon bemerkt haben, als Sie zum ersten Mal in einem Basar feilschten, hat der anfängliche Anker einen starken Effekt. Als ich meine Studenten in der Kunst des Verhandelns unterrichtete, riet ich ihnen, dass sie im Fall, dass die andere Seite ein unverschämtes Angebot gemacht hat, nicht mit einem ebenso unverschämten Gegenangebot kontern sollten,

das eine in zukünftigen Verhandlungen schwer zu überbrückende Kluft schaffen würde. Stattdessen sollten sie eine Show abziehen, aus dem Zimmer stürmen oder damit drohen und – sich selbst und der anderen Seite – unmissverständlich klarmachen, dass sie die Verhandlungen mit dieser Zahl auf dem Tisch nicht fortsetzen würden.

Die Psychologen Adam Galinsky und Thomas Mussweiler schlugen subtilere Strategien vor, um bei Verhandlungen dem Ankereffekt zu widerstehen. [9] Sie wiesen die Verhandlungsführer an, ihre Aufmerksamkeit zu fokussieren und ihr Gedächtnis nach Argumenten gegen den Anker zu durchforsten. Die Anweisung, System 2 zu aktivieren, war erfolgreich. So wird der Ankereffekt zum Beispiel reduziert oder auch ganz aufgehoben, wenn derjenige, der den zweiten Zug macht, seine Aufmerksamkeit auf das Mindestangebot lenkt, das der Gegner akzeptieren würde, oder auf die Kosten, die dem Gegner dadurch entstehen, dass er keine Einigung erzielt. Grundsätzlich mag eine Strategie des bewussten »Denkens des Gegenteils« eine gute Verteidigung gegen Ankereffekte sein, weil sie die verzerrte Rekrutierung von Gedanken, die diese Effekte erzeugen, beseitigt.

Versuchen Sie schließlich einmal, den Ankereffekt bei einem Problem von öffentlichem Interesse herauszuarbeiten: der Höhe des Schadensersatzes in Fällen von Körperverletzung. Hier werden den Geschädigten mitunter sehr hohe Summen zugesprochen. Unternehmen, gegen die sich solche Klagen vielfach richten, wie Kliniken und Chemiefirmen, haben sich für gesetzliche Obergrenzen von Haftungssummen starkgemacht. Bevor Sie dieses Kapitel gelesen haben, haben Sie vielleicht geglaubt, dass Entschädigungsobergrenzen für potenzielle Beklagte zweifellos eine gute Sache seien, aber jetzt sollten Sie da nicht mehr so sicher sein. Betrachten Sie den Effekt einer Kappung von Schadensersatzansprüchen bei einer Million Dollar. Diese Regel würde dafür sorgen, dass es keine höheren Entschädigungssummen mehr gibt, aber zugleich würde der Anker viele Entschädigungsbeträge, die ansonsten viel kleiner wären, in die Höhe treiben.[10] Sie würde höchstwahrscheinlich Täter, die sich einer schweren Körperverletzung schuldig gemacht hätten, und Großunternehmen viel stärker begünstigen als kleinere Firmen.

## Ankerung und die beiden Systeme

Die Wirkungen von zufällig ausgewählten Ankern sagen uns eine Menge über die Beziehung zwischen System 1 und System 2. Ankereffekte wurden immer bei Urteils- und Entscheidungsaufgaben erforscht, die letztlich von System 2 abgeschlossen werden. Doch System 2 arbeitet mit Daten, die im Rahmen einer automatischen und unwillkürlichen Operation von System 1 aus dem Gedächtnis abgerufen werden. Daher ist System 2 anfällig für den verzerrenden Einfluss von Ankern, die bestimmte Informationen leichter abrufbar machen. Außerdem hat System 2 keine Kontrolle über den Effekt und weiß nichts davon. Die Versuchsteilnehmer, denen zufällig ausgewählte oder absurde Anker (wie etwa Gandhis Tod mit 144 Jahren) dargeboten wurden, bestreiten entschieden, dass diese offensichtlich wertlose Information ihre Schätzung beeinflusst haben könnte – aber sie irren sich.

Wir sahen in der Diskussion des Gesetzes der kleinen Zahlen, dass eine Nachricht, sofern sie nicht sofort als Lüge verworfen wird, unabhängig von ihrer Vertrauenswürdigkeit die gleiche Wirkung auf das assoziative System haben wird. Der Kern der Nachricht besteht aus der Geschichte, die auf der jeweils verfügbaren Information basiert, auch wenn die Informationsmenge gering und ihre Qualität schlecht ist: WYSIATI. Wenn wir eine Geschichte über die heldenhafte Rettung eines verwundeten Bergsteigers lesen, ist ihre Wirkung auf unser assoziatives Gedächtnis weitgehend unabhängig davon, ob es sich um eine Nachrichtenmeldung oder die Zusammenfassung eines Films handelt. Die Ankerung ist das Ergebnis dieser assoziativen Aktivierung. Ob die Geschichte wahr ist oder auch nur glaubwürdig, spielt keine große Rolle, wenn überhaupt. Die Wirkmächtigkeit von Zufallsankern ist ein Extremfall dieses Phänomens, weil ein Zufallsanker offensichtlich keinerlei Information liefert.

Weiter vorn erörterte ich die verwirrende Vielfalt von Priming-Effekten, bei denen unser Denken und Verhalten von Stimuli beeinflusst werden kann, denen man keine Aufmerksamkeit schenkt, und sogar von Stimuli, die gänzlich unbemerkt bleiben. Die wichtigste Erkenntnis aus der Priming-Forschung besteht darin, dass unser Denken und unser Verhalten in viel stärkerem Maße, als wir es erkennen oder wollen, von unserem augenblicklichen Umfeld beeinflusst wird. Viele Menschen bezweifeln die

Ergebnisse von Priming-Experimenten, weil sie nicht ihrem subjektiven Erleben entsprechen. Viele andere finden die Ergebnisse erschütternd, weil sie darin eine Bedrohung für das subjektive Bewusstsein von Handlungskompetenz und autonomer Selbstbestimmung sehen. Wenn der Inhalt eines Bildschirmschoners auf einem belanglosen Computer unsere Bereitschaft, Fremden zu helfen, beeinflussen kann, ohne dass wir uns dessen bewusst sind, wie frei sind wir dann? Ankereffekte werden als ähnlich bedrohlich erlebt. Obwohl wir uns des Ankers immer bewusst sind und ihm sogar unsere Aufmerksamkeit schenken, wissen wir nicht, wie er unser Denken lenkt und einschränkt, weil wir uns nicht vorstellen können, was wir gedacht hätten, wenn der Anker anders gewesen wäre (oder gar kein Anker vorhanden gewesen wäre). Wir sollten jedenfalls davon ausgehen, dass jede Zahl, die uns dargeboten wird, einen Ankereffekt auf uns hat, und wenn viel auf dem Spiel steht, sollte man sich selbst (sein System 2) mobilisieren, um diesen Effekt zu bekämpfen.

## Zum Thema »Anker«

»Die Firma, die wir übernehmen wollen, ließ uns ihren Geschäftsplan mit dem erwarteten Umsatz zukommen. Wir sollten uns nicht von dieser Zahl beeinflussen lassen. Wir sollten sie ignorieren.«

»Pläne sind optimale Szenarien. Bei der Vorhersage der tatsächlichen Ergebnisse sollten wir uns nicht von dem Ankereffekt der Planzahlen beeinflussen lassen. Dies können wir unter anderem dadurch tun, dass wir uns überlegen, was bei dem Plan schiefgehen könnte.«

»Unser Ziel bei den Verhandlungen ist es, ihnen diese Zahl als Anker unterzujubeln.«

»Stellen wir klar, dass wir die Verhandlungen abbrechen, wenn dies ihr Angebot ist. Das ist für uns keine akzeptable Verhandlungsbasis.«

»Die Anwälte des Beklagten erwähnten scheinbar beiläufig einen Fall, in dem eine lächerlich geringe Schadensersatzsumme zuerkannt wurde. Und es gelang ihnen so, diesen Betrag als Anker für die Entscheidung des Richters festzusetzen.« [11]

# *12. Die Wissenschaft der Verfügbarkeit*

Die produktivsten Jahre der Zusammenarbeit zwischen Amos und mir waren 1971 und 1972. Wir waren damals Gäste des Oregon Research Institute in Eugene, das mehrere zukünftige Stars sämtlicher Forschungsfelder, auf denen wir arbeiteten, beherbergte – Urteils- und Entscheidungsfindung und intuitive Vorhersagen. Unser wichtigster Gastgeber war Paul Slovic, der in Ann Arbor mit Amos studiert hatte und lebenslang mit ihm befreundet blieb. Paul stand im Begriff, der führende Psychologe auf dem Gebiet der Risikoforschung zu werden – eine Position, die er seit Jahrzehnten innehat und die ihm zahllose Ehrungen eintrug. Paul und seine Frau Roz machten uns mit dem Lebensstil in Eugene vertraut, und bald taten wir das, was alle Menschen in Eugene tun – joggen, grillen und Kinder zu Basketballspielen mitnehmen. Aber wir arbeiteten auch sehr hart, führten Dutzende von Experimenten durch und verfassten unsere Artikel über Urteilsheuristiken. Abends schrieb ich an *Attention and Effort*. Es war ein arbeitsreiches Jahr.

Eines unserer Projekte war die Erforschung eines Phänomens, das wir »Verfügbarkeitsheuristik« nannten. Wir kamen auf diese Heuristik, als wir uns fragten, was Menschen tun, wenn sie die Häufigkeit einer bestimmten Kategorie abschätzen wollen, wie etwa »Menschen, die sich nach dem sechzigsten Lebensjahr scheiden lassen« oder »gefährliche Fabriken«. Die Antwort war einfach: Beispiele der jeweiligen Kategorien werden aus dem Gedächtnis abgerufen, und wenn der Abruf leicht und flüssig ist, wird die Kategorie als groß beurteilt. Wir definierten die Verfügbarkeitsheuristik als den Prozess der Einschätzung von Häufigkeit »anhand der Leichtigkeit, mit der Beispielfälle erinnert werden«. [1] Diese Aussage schien klar und deutlich zu sein, als wir sie formulierten, aber der Verfügbarkeitsbegriff wurde seither weiter verfeinert.

Als wir die Verfügbarkeitsheuristik erforschten, war das Modell der zwei Systeme noch nicht entwickelt worden, und wir versuchten nicht, herauszufinden, ob diese Heuristik eine gezielte Strategie der Problemlösung oder eine automatische Operation ist. Heute wissen wir, dass beide Systeme beteiligt sind.

Wir haben uns frühzeitig mit der Frage befasst, wie viele Beispielfälle abgerufen werden müssen, um einen Eindruck von der Leichtigkeit zu bekommen, mit der sie erinnert werden. Wir wissen heute die Antwort: keiner. Betrachten Sie die beiden nachstehenden Buchstabenfolgen, und überlegen Sie, wie viele Wörter jeweils daraus gebildet werden können.

XUZONLCJM
TAPCERHOB

Sie wussten praktisch auf Anhieb – ohne Beispielwörter zu bilden –, dass eine Folge mehr Möglichkeiten bietet als die andere, vermutlich mindestens zehnmal mehr Möglichkeiten. In ähnlicher Weise müssen Sie keine konkreten Nachrichtenmeldungen abrufen, um eine gute Vorstellung von der relativen Häufigkeit zu bekommen, mit der verschiedene Länder im letzten Jahr in den Nachrichten auftauchten (Belgien, China, Frankreich, Kongo, Nicaragua, Rumänien …).

Wie andere Urteilsheuristiken ersetzt auch die Verfügbarkeitsheuristik eine Frage durch eine andere: Sie wollen die Größe einer Kategorie oder die Häufigkeit eines Ereignisses abschätzen, aber Sie berichten darüber, wie leicht Ihnen Beispielfälle eingefallen sind. Die Ersetzung von Fragen erzeugt zwangsläufig systematische Fehler. Sie können selbst entdecken, wie die Heuristik zu Verzerrungen führt, indem Sie einer einfachen Vorgehensweise folgen: Listen Sie andere Faktoren als die Häufigkeit auf, die es Ihnen erleichtern, Beispiele zu generieren. Jeder Faktor auf Ihrer Liste ist eine potenzielle Ursache für Verzerrungen. Hier einige Beispiele:

– Ein hervorstechendes (salientes) Ereignis, das Ihre Aufmerksamkeit auf sich zieht, lässt sich leicht aus dem Gedächtnis abrufen. Scheidungen zwischen Hollywoodstars und Sexskandale von Politikern erregen viel Aufmerksamkeit, und Beispiele fallen Ihnen schnell ein. Daher werden Sie wahrscheinlich die Häufigkeit von Scheidungen in Hollywood und Sexskandalen bei Politikern überschätzen.
– Ein dramatisches Ereignis erhöht vorübergehend die Verfügbarkeit seiner Kategorie. Ein Flugzeugabsturz, über den in den Medien berichtet wird, verändert kurzzeitig Ihre gefühlsmäßige Einstellung zur

Sicherheit des Fliegens. Nachdem Sie ein brennendes Auto am Straßenrand gesehen haben, sind Ihnen Unfälle mental präsent, und die Welt ist eine Zeit lang ein gefährlicherer Ort.

– Persönliche Erfahrungen, Bilder und anschauliche Beispiele sind verfügbarer als Begebenheiten, die anderen widerfahren sind, bloße Wörter oder Statistiken. Ein gerichtliches Fehlurteil, von dem Sie betroffen sind, untergräbt Ihr Vertrauen in das Justizsystem stärker als ein ähnliches Ereignis, über das Sie in der Zeitung lesen.

Es ist möglich, aber anstrengend, dieser großen Zahl potenzieller Verfügbarkeitsfehler zu widerstehen. Man muss seine Eindrücke und Intuitionen mühsam überdenken, indem man Fragen stellt wie: »Ist unsere Einschätzung, dass Diebstähle durch Teenager ein großes Problem sind, nur darauf zurückführen, dass es in letzter Zeit in unserem Viertel einige Fälle gegeben hat?« oder »Könnte es sein, dass ich das Gefühl habe, mich nicht gegen Grippe impfen lassen zu müssen, weil keiner meiner Bekannten im letzten Jahr an Grippe erkrankte?« . Vor solchen Verzerrungen auf der Hut zu sein ist anstrengend – aber die Chance, einen kostspieligen Fehler zu vermeiden, ist manchmal die Mühe wert.

Eine der bekanntesten Studien über Verfügbarkeit deutet darauf hin, dass es der Harmonie in Beziehungen und in anderen gemeinsamen Projekten zuträglich ist, wenn man sich seiner kognitiven Verzerrungen bewusst ist. In einer berühmten Studie wurden Eheleute gefragt: »Wie groß war Ihr persönlicher Beitrag beim Aufräumen Ihrer gemeinsamen Wohnung in Prozent?« Sie beantworteten auch ähnliche Fragen über »das Beseitigen des Mülls«, die »Anbahnung sozialer Kontakte« und so weiter. Würden sich die selbst geschätzten Beiträge zu 100 Prozent – oder mehr, oder weniger – addieren?[2] Wie erwartet, addierten sich die Beiträge auf über 100 Prozent. Die Erklärung ist ein einfacher Verfügbarkeitsfehler: Beide Eheleute erinnern sich an ihre eigenen Bemühungen und Beiträge sehr viel deutlicher als an die ihres jeweiligen Partners, und dieser Verfügbarkeitsunterschied führt zu einem Unterschied in der beurteilten Häufigkeit. Diese Verzerrung ist nicht unbedingt eigennützig: Eheleute überschätzten auch ihre Beiträge zu Streitigkeiten, wenn auch in einem geringeren Ausmaß als ihre Beiträge zu erwünschten Ergebnissen. Die gleiche Verzerrung trägt zu der weitverbreiteten Beobachtung bei, dass viele Mitglieder eines Teams den Eindruck haben, mehr getan zu haben, als eigentlich angemessen gewesen

wäre, und sie haben auch das Gefühl, dass die anderen ihnen für ihre individuellen Beiträge nicht hinreichend dankbar sind.

Grundsätzlich bin ich nicht optimistisch, was unsere Fähigkeit zur Kontrolle von Verzerrungen anbelangt, aber dies ist eine Ausnahme. Die Chance zur erfolgreichen Aufhebung von Verzerrungen existiert, weil die Umstände, unter denen Streitigkeiten über die angemessene Würdigung von Beiträgen auftreten, leicht zu identifizieren sind, und dies umso mehr, weil Spannungen oftmals dann auftreten, wenn mehrere Personen gleichzeitig das Gefühl haben, dass ihre Anstrengungen nicht gebührend gewürdigt werden. Die bloße Feststellung, dass normalerweise von allen zusammen über 100 Prozent Anerkennung beansprucht wird, genügt manchmal, um die Situation zu entschärfen. Jedenfalls sollte sich dies jeder Einzelne in Erinnerung rufen. Man leistet gelegentlich einen größeren Beitrag, als es angemessen wäre, aber es ist hilfreich, zu wissen, dass man dieses Gefühl vermutlich auch dann haben wird, wenn alle anderen Mitglieder des Teams das gleiche Gefühl haben.

## Die Psychologie der Verfügbarkeit

Ein großer Fortschritt beim Verständnis der Verfügbarkeitsheuristik gelang Anfang der 1990er-Jahre, als eine Gruppe deutscher Psychologen unter Leitung von Norbert Schwarz eine faszinierende Frage aufwarf: Wie werden unsere Eindrücke von der Häufigkeit einer Kategorie durch die Aufgabe beeinflusst, eine vorgegebene Anzahl von Beispielen aufzulisten?[3]

Stellen Sie sich vor, Sie wären selbst eine Versuchsperson in diesem Experiment:

> Listen Sie zunächst sechs Situationen auf, in denen Sie sich durchsetzungsfähig verhalten haben.
> Beurteilen Sie dann, wie durchsetzungsfähig Sie waren.

Stellen Sie sich vor, man hätte Sie nach zwölf Beispielen von durchsetzungsfähigem Verhalten gefragt (eine Zahl, die den meisten Menschen Mühe macht). Würden Sie Ihr Durchsetzungsvermögen nun anders beurteilen?

Schwarz und seine Kollegen stellten fest, dass die Aufgabe, Beispiele aufzulisten, die Beurteilungen dieser Eigenschaft über zwei verschiedene Wege verbessern kann:

– Die Anzahl der abgerufenen Beispiele
– Die Leichtigkeit des Abrufs

Die Forderung, zwölf Beispiele aufzulisten, spielt die beiden Determinanten gegeneinander aus. Einerseits haben Sie sich gerade eine erstaunliche Zahl von Fällen, in denen Sie Durchsetzungsvermögen gezeigt haben, in Erinnerung gerufen. Andererseits ist es so, dass Ihnen zwar die ersten drei oder vier Beispiele für Ihre Durchsetzungsfähigkeit vermutlich leicht eingefallen sind, aber dann haben Sie sich mit Sicherheit ziemlich abgemüht, um die Zwölferliste zu vervollständigen; die Abrufflüssigkeit

*(fluency)* war gering. Was zählt mehr – die Anzahl der abgerufenen Beispiele oder die Leichtigkeit und Flüssigkeit des Abrufs?

Das Experiment erbrachte einen eindeutigen Gewinner: Menschen, die gerade zwölf Beispiele aufgelistet hatten, beurteilten sich selbst als weniger durchsetzungsfähig als Personen, die nur sechs Beispiele angeführt hatten. Außerdem hielten sich Probanden, die zwölf Situationen auflisten sollten, in denen sie sich *nicht* durchsetzungsfähig verhalten hatten, schließlich für ziemlich durchsetzungsstark! Wenn einem nicht mühelos Situationen einfallen, in denen man klein beigegeben hat, gelangt man wahrscheinlich zu dem Schluss, dass man alles andere als kleinlaut ist. Selbsteinschätzungen richteten sich maßgeblich nach der Leichtigkeit, mit der den Probanden Beispiele einfielen. Das Erleben des flüssigen Abrufens von Beispielen war wichtiger als die Anzahl der abgerufenen Beispiele.

Ein noch direkterer Nachweis der Rolle der Abrufflüssigkeit lieferten andere Psychologen derselben Forschergruppe.[4] Alle Teilnehmer an ihrem Experiment listeten sechs Beispiele von durchsetzungsfähigem (oder nicht durchsetzungsfähigem) Verhalten auf, während sie einen bestimmten Gesichtsausdruck aufrechterhalten sollten. »Lächler« sollten den Jochbeinmuskel zusammenziehen, wodurch ein leichtes Lächeln entsteht; »Stirnrunzler« sollten ihre Stirn in Falten legen. Wie Sie bereits wissen, geht Stirnrunzeln normalerweise mit kognitiver Beanspruchung einher, und der Effekt ist symmetrisch: Wenn Probanden die Stirn runzeln sollen, während sie eine Aufgabe erledigen, strengen sie sich stärker an und erleben eine größere kognitive Beanspruchung. Die Forscher erwarteten, dass es den Stirnrunzlern schwererfallen würde, Beispiele von durchsetzungsfähigem Verhalten abzurufen, und dass sie sich daher als vergleichsweise wenig durchsetzungsfähig einstufen würden. Und das war auch der Fall.

Psychologen gefallen Experimente, die paradoxe Ergebnisse bringen, und sie haben Schwarz' Entdeckung voller Begeisterung angewandt. Hier einige ihrer Befunde:

– Menschen glauben, dass sie ihr Fahrrad weniger oft benutzen, nachdem sie sich an viele statt an wenige Fälle des Fahrradgebrauchs erinnern.
– Menschen sind weniger überzeugt von der Richtigkeit einer Wahl, wenn sie mehr Pro-Argumente dafür anführen sollen.

– Menschen sind weniger überzeugt davon, dass ein Ereignis vermeidbar war, nachdem sie mehr Möglichkeiten, es zu vermeiden, aufgelistet haben.
– Menschen sind weniger beeindruckt von einem Auto, nachdem sie viele seiner Vorteile aufgelistet haben.

Ein Professor an der Universität von Kalifornien in Los Angeles fand eine geniale Methode, um den Verfügbarkeitsfehler auszunutzen. Er bat verschiedene Gruppen von Studenten, Verbesserungsvorschläge für seine Lehrveranstaltungen aufzulisten, und er variierte die geforderte Zahl von Verbesserungsvorschlägen. Erwartungsgemäß beurteilten die Studenten, die mehr Verbesserungsvorschläge machten, seinen Unterricht besser!

Der vielleicht interessanteste Befund dieser paradoxen Forschungen ist die Tatsache, dass das Paradoxon nicht immer auftritt: Menschen orientieren sich manchmal am Inhalt statt an der Abrufleichtigkeit. Man versteht ein Verhaltensmuster erst dann richtig, wenn man weiß, wie man es umkehren kann. Schwarz und seine Kollegen stellten sich dieser Herausforderung, die Bedingungen zu entdecken, unter denen diese Umkehr stattfinden würde.

Die Leichtigkeit, mit der den Probanden Beispiele von Durchsetzungsvermögen einfallen, verändert sich im Lauf der Aufgabe. Die ersten Beispiele sind leicht, aber die Abrufung wird schon bald viel schwerer. Selbstverständlich erwartet die Versuchsperson auch, dass die Flüssigkeit allmählich zurückgeht, aber die Flüssigkeit scheint zwischen sechs und zwölf Beispielen stärker abzufallen, als die Versuchsperson vermutete. Die Ergebnisse deuten darauf hin, dass die Probanden eine Schlussfolgerung ziehen: Wenn es mir so viel schwererfällt, als ich erwartet habe, Beispiele für meine Durchsetzungsfähigkeit zu finden, kann ich nicht besonders durchsetzungsfähig sein. Beachten Sie, dass diese Schlussfolgerung auf einer Überraschung basiert – die Abrufflüssigkeit ist schlechter als erwartet. Die Verfügbarkeitsheuristik, die die Probanden anwenden, sollte besser »unerklärte Unverfügbarkeitsheuristik« genannt werden.

Schwarz und seine Mitarbeiter folgerten, dass sie die Heuristik dadurch stören könnten, dass sie den Probanden eine Erklärung für die Abrufflüssigkeit, die sie erlebten, gaben. Sie sagten den Versuchspersonen, sie würden Hintergrundmusik hören, während sie Beispiele für Durchsetzungsvermögen abrufen, und die Musik werde ihre

Leistungsfähigkeit bei der Gedächtnisaufgabe beeinflussen. Einigen Probanden wurde gesagt, die Musik werde ihnen helfen, anderen wurde gesagt, sie sollten mit einer verminderten Flüssigkeit rechnen. Wie vorhergesagt, benutzten die Probanden, deren erlebte Abrufflüssigkeit »erklärt« wurde, diese nicht als Heuristik; die Versuchsteilnehmer, denen gesagt wurde, Musik erschwere den Abrufprozess, schätzten sich selbst als gleich durchsetzungsfähig ein, ganz egal ob sie zwölf oder sechs Beispiele abriefen. Andere »Legenden« wurden mit dem gleichen Resultat benutzt: Urteile werden nicht länger von der Abrufleichtigkeit beeinflusst, wenn für das subjektive Erleben der Flüssigkeit eine Pseudoerklärung gegeben wird, entweder durch die Präsenz kurvenförmiger oder gerader Textfelder, durch die Hintergrundfarbe des Bildschirms oder durch andere irrelevante Faktoren, die sich die Experimentatoren ausgedacht haben.[5]

So, wie ich ihn beschrieben habe, scheint der Prozess, der zu Verfügbarkeitsurteilen führt, mit einem komplexen Gedankengang verbunden zu sein. Die Probanden erleben eine verminderte Flüssigkeit beim Abrufen von Beispielen. Sie haben offensichtlich Erwartungen hinsichtlich der Schnelligkeit, mit der die Flüssigkeit abnimmt, und diese Erwartungen sind falsch: Die Schwierigkeit, neue Beispiele zu finden, wächst schneller, als sie es erwarten. Die unerwartet niedrige Abrufflüssigkeit veranlasst Menschen, die um zwölf Beispiele gebeten wurden, dazu, sich als nicht durchsetzungsfähig zu beschreiben. Wenn die Überraschung beseitigt wird, wird das Urteil nicht länger von der niedrigen Flüssigkeit beeinflusst. Der Prozess scheint aus einer komplexen Reihe von Schlussfolgerungen zu bestehen. Ist das automatische System 1 dazu in der Lage?

Die Antwort lautet, dass tatsächlich kein komplexer Gedankengang erforderlich ist. Zu den Grundmerkmalen von System 1 gehört die Fähigkeit, Erwartungen festzulegen und überrascht zu sein, wenn diese Erwartungen nicht erfüllt werden. Das System ruft auch potenzielle Ursachen einer Überraschung ab, für gewöhnlich dadurch, dass es unter den jüngsten Überraschungen eine mögliche Ursache identifiziert. Außerdem kann System 2 die Erwartungen von System 1 spontan neu einstellen, sodass ein Ereignis, das normalerweise überraschend wäre, jetzt beinahe normal ist. Angenommen, man sagt Ihnen, dass ein dreijähriger Junge aus Ihrer Nachbarschaft häufig einen Zylinder trägt, wenn er im Kinderwagen sitzt. Wenn Sie ihn dann tatsächlich mit diesem Zylinder sehen, wird Sie

das viel weniger überraschen, als wenn Sie nicht vorgewarnt worden wären. In Schwarz' Experiment wurde die Hintergrundmusik als eine mögliche Ursache von Abrufproblemen erwähnt. Die Schwierigkeit, sich an zwölf Beispiele zu erinnern, ist keine Überraschung mehr, und daher ist es weniger wahrscheinlich, dass sie durch die Aufgabe, die eigene Durchsetzungsfähigkeit zu beurteilen, hervorgerufen wird.

Schwarz und seine Kollegen fanden heraus, dass Menschen, die ein persönliches Interesse an dem Urteil haben, eher die Anzahl der abgerufenen Beispiele betrachten und sich weniger nach der Flüssigkeit richten. Im Zuge einer Studie über die Risiken für Herzerkrankungen stellten sie zwei Studentengruppen zusammen. Die Hälfte der Studenten hatten Verwandte mit Herzerkrankungen, und daher wurde von ihnen erwartet, dass sie die Aufgabe ernster nehmen als die anderen, die in ihrer Familie keine Herzkranken hatten. Alle sollten entweder drei oder acht gewohnheitsmäßige Verhaltensweisen anführen, die sich auf ihre Herzgesundheit auswirken könnten (einige wurden nach riskanten Gewohnheiten gefragt, andere nach schützenden Gewohnheiten).[6] Studenten, die familiär nicht durch Herzkrankheiten vorbelastet waren, gingen salopp an die Aufgabe heran und hielten sich an die Verfügbarkeitsheuristik. Studenten, die Mühe hatten, acht Beispiele für riskantes Verhalten zu finden, fühlten sich selbst vergleichsweise sicher, und diejenigen, die mühsam nach Beispielen von protektiven Verhaltensweisen suchten, stuften sich als gefährdet ein. Studenten mit familiärer Vorbelastung zeigten das entgegengesetzte Muster – sie fühlten sich sicherer, wenn ihnen viele Beispiele von sicherem Verhalten einfielen, und sie fühlten eine größere Bedrohung, wenn sie sich an viele Beispiele riskanten Verhaltens erinnerten. Sie waren auch eher der Auffassung, dass die Erfahrung der Risikobewertung sich auf ihr zukünftiges Verhalten auswirken würde.

Daraus lässt sich folgern, dass die Abrufleichtigkeit eine System-1-Heuristik ist, die durch einen Inhaltsfokus ersetzt wird, wenn System 2 stärker beansprucht wird. Die Ergebnisse vielfältiger Studien legen die Schlussfolgerung nahe, dass Menschen, die sich von System 1 lenken lassen, anfälliger für Verfügbarkeitsfehler sind als andere, die sich in einem Zustand erhöhter Vigilanz befinden. Nachfolgend sind einige Zustände aufgelistet, in denen Menschen »mit dem Strom schwimmen« und stärker

von der Abrufleichtigkeit als von dem abgerufenen Inhalt beeinflusst werden:

- Wenn sie gleichzeitig mit einer anderen anstrengenden Aufgabe beschäftigt sind.[7]
- Wenn sie gut gelaunt sind, weil sie gerade an eine glückliche Episode in ihrem Leben gedacht haben.[8]
- Wenn sie auf einer Depressionsskala niedrige Werte aufweisen.[9]
- Wenn sie sachkundige Neulinge bei dem Thema der Aufgabe sind,[10] im Unterschied zu wahren Experten.[11]
- Wenn sie hohes Vertrauen in ihre Intuition haben.[12]
- Wenn sie Macht haben (oder man ihnen das Gefühl gibt, Macht zu haben). [13]

Den letzten Befund finde ich besonders bemerkenswert. Die Autoren beginnen ihren Artikel mit einem berühmten Zitat: »Ich verbringe nicht viel Zeit damit, Meinungsumfragen rund um die Welt durchzuführen, um herauszufinden, ob das, was ich glaube, richtig ist. Ich muss nur wissen, was ich fühle.« (George W. Bush, November 2002). Sie zeigen dann, dass Vertrauen auf Intuition nur zum Teil ein Persönlichkeitszug ist. Es genügt schon, Menschen daran zu erinnern, dass sie einmal Macht besessen haben, um ihr scheinbares Vertrauen in ihre Intuition zu steigern.

## Zum Thema »Verfügbarkeit«

> »Weil letzten Monat zufälligerweise zwei Flugzeuge abgestürzt sind, nimmt sie jetzt lieber den Zug. Das ist dumm. An dem Risiko hat sich nichts geändert; sie macht einen Verfügbarkeitsfehler.«

»Er unterschätzt die Risiken durch Luftschadstoffe in Innenräumen, weil die Medien kaum darüber berichten. Das ist ein Verfügbarkeitseffekt. Er sollte sich die Statistiken ansehen.«

»Sie hat in letzter Zeit zu viele Spionagefilme gesehen, denn jetzt sieht sie überall Verschwörungen.«

»Die Vorstandschefin hatte mehrere Erfolge hintereinander, sodass ihr Misserfolge nicht so ohne Weiteres einfallen. Der Verfügbarkeitsfehler macht sie übertrieben optimistisch.«

# *13. Verfügbarkeit, Emotion und Risiko*

Risikoforscher haben schnell erkannt, dass das Konzept der Verfügbarkeit auch für sie von Interesse ist. Bereits vor der Publikation unserer Arbeit bemerkte der Ökonom Howard Kunreuther, der damals am Anfang einer wissenschaftlichen Laufbahn stand, die er der Erforschung von Risiken und Versicherungen widmete, dass Verfügbarkeitseffekte helfen, das Muster von Versicherungsabschlüssen und Schutzmaßnahmen nach Katastrophen zu erklären. Nach einer Katastrophe sind Schadensopfer und Beinahe-Opfer sehr besorgt. Nach jedem größeren Erdbeben schließen die Kalifornier eine Zeit lang fleißig Versicherungen ab und ergreifen Maßnahmen zum Schutz ihrer Gebäude und zur Schadensminderung. Sie zurren ihre Heizkessel fest, um die Schäden durch ein Erdbeben zu verringern, dichten ihre Kellertüren gegen Überschwemmungen ab und legen in ausreichendem Maße Notvorräte an. Doch die Erinnerungen an die Katastrophe verblassen mit der Zeit, und mit der Sorge nimmt auch die Vorsorge ab. Die Dynamik des Gedächtnisses hilft, die wiederkehrenden Zyklen von Katastrophe, Risikoschutz und zunehmender Sorglosigkeit zu erklären, die all jenen vertraut sind, die große Katastrophen erforschen.

Kunreuther stellte auch fest, dass Schutzmaßnahmen von Privatpersonen oder von Staaten in der Regel auf den schwersten Katastrophenfall, der sich bislang ereignet hat, zugeschnitten sind. Schon die alten Ägypter haben die Hochwassermarken von Flüssen, die in regelmäßigen Abständen über die Ufer traten, aufmerksam registriert – und sich immer entsprechend gewappnet, ganz offensichtlich in der Annahme, dass die Fluten nicht über die aktuelle Hochwassermarke hinaus steigen werden. Bilder einer schlimmeren Katastrophe kommen uns nicht ohne Weiteres in den Sinn.

## Verfügbarkeit und Affekt

Die einflussreichsten Studien über Verfügbarkeitsfehler wurden von unseren Freunden in Eugene durchgeführt, wo sich unser ehemaliger Student Baruch Fischhoff Paul Slovic und dessen langjähriger Mitarbeiterin Sarah Lichtenstein anschloss. Sie führten bahnbrechende Forschungen über öffentliche Wahrnehmungen von Risiken durch, unter anderem auch eine Umfrage, die zum Standardbeispiel für einen Verfügbarkeitsfehler geworden ist. Den Teilnehmern der Studie wurden jeweils Paare von Todesursachen gezeigt: Diabetes und Asthma oder Schlaganfall und Unfälle. Bei jedem Paar sollten die Probanden die häufigere Ursache angeben und das Verhältnis der beiden Häufigkeiten abschätzen. Die Urteile wurden mit aktuellen Gesundheitsstatistiken verglichen.

Hier sind einige ihrer Ergebnisse:

- Schlaganfälle verursachen fast doppelt so viele Todesfälle wie alle Unfälle zusammengenommen, aber 80 Prozent der Probanden stuften einen Unfalltod als wahrscheinlicher ein.
- Tornados sollten mehr Menschen das Leben kosten als Asthma, obwohl Asthma zwanzigmal mehr Todesopfer fordert.
- Tod durch Blitzschlag wurde als weniger wahrscheinlich eingestuft als Tod durch Lebensmittelvergiftung, obwohl er 52-mal häufiger ist.
- Krankheiten fordern etwa 18-mal mehr Todesopfer als Unfälle, aber beide Todesursachen wurden als etwa gleich wahrscheinlich beurteilt.
- Unfälle sollten über 300-mal mehr Todesopfer fordern als Diabetes, während das Verhältnis in Wirklichkeit 1:4 beträgt.

Die Lektion ist klar: Einschätzungen von Todesursachen werden durch die Medienberichterstattung verzerrt. Die Berichterstattung ihrerseits wird durch Neuigkeit und Prägnanz verzerrt. Die Medien prägen nicht nur das Interesse der Öffentlichkeit, sie werden ihrerseits von diesem Interesse beeinflusst. Redakteure können die Forderungen der Öffentlichkeit nach ausführlicher Berichterstattung über bestimmte Themen und Standpunkte nicht ignorieren. Ungewöhnliche Ereignisse (etwa Lebensmittelvergiftungen) ziehen unverhältnismäßig viel Aufmerksamkeit auf sich und werden daher als weniger ungewöhnlich wahrgenommen, als

sie es tatsächlich sind. Die Welt in unseren Köpfen ist keine exakte Kopie der Wirklichkeit; unsere Erwartungen bezüglich der Häufigkeit von Ereignissen werden durch die Verbreitung und emotionale Intensität der Nachrichten, denen wir ausgesetzt sind, verzerrt.

In den Schätzungen von Todesursachen spiegelt sich die Aktivierung von Vorstellungen im assoziativen Gedächtnis praktisch direkt wider, und sie sind ein gutes Beispiel für Ersetzungen. Aber Slovic und seine Kollegen gelangten zu einer tieferen Einsicht: Sie erkannten, dass die Leichtigkeit, mit der uns Vorstellungen verschiedener Risiken einfallen, und die emotionalen Reaktionen auf diese Risiken untrennbar miteinander verbunden sind. Erschreckende Gedanken und Bilder fallen uns besonders leicht ein, und leicht abrufbare und anschauliche Vorstellungen von Gefahren verstärken die Furcht.

Wie bereits erwähnt, erarbeitete Slovic schließlich das Konzept der Affektheuristik, das besagt, dass wir bei Urteilen und Entscheidungen unsere Emotionen zurate ziehen: Mag ich das? Hasse ich es? Wie stark reagiere ich emotional darauf? In vielen Lebensbereichen, so Slovic, bilden sich Menschen Meinungen und treffen Entscheidungen, die unmittelbar ihre Gefühle und ihre grundlegende Annäherungs- oder Vermeidungstendenz zum Ausdruck bringen, oftmals ohne zu wissen, dass sie dies tun. Die Affektheuristik ist ein Fall von Ersetzung, bei dem die Antwort auf eine leichte Frage (Welche Gefühle weckt das in mir?) als Antwort auf eine viel schwierigere Frage dient (Was denke ich darüber?). Slovic und seine Mitarbeiter stellten ihre Erkenntnisse in den Kontext der Arbeiten des Neurowissenschaftlers Antonio Damasio, der die Ansicht vertritt, dass unsere emotionalen Bewertungen von Ergebnissen und die damit verbundenen Körperzustände und Annäherungs- sowie Vermeidungstendenzen eine zentrale Rolle bei der menschlichen Entscheidungsfindung spielen. Damasio und seine Mitarbeiter haben beobachtet, dass Menschen, die, manchmal aufgrund von Hirnschädigungen, nicht die angemessenen Emotionen zeigen, bevor sie eine Entscheidung treffen, auch in ihrer Fähigkeit beeinträchtigt sind, gute Entscheidungen zu treffen.[1] Die Unfähigkeit, sich von einer »gesunden Furcht« vor negativen Konsequenzen leiten zu lassen, ist ein katastrophaler Defekt.

Slovics Forschungsgruppe hat die Funktionsweise der Affektheuristik in einer Studie sehr überzeugend nachvollzogen. Sie befragten Probanden

nach ihren Meinungen über verschiedene Technologien wie Trinkwasser-Fluoridierung, Chemiefabriken, Konservierungsstoffe in Lebensmitteln und Autos, und sie forderten die Teilnehmer auf, den Nutzen und die Risiken jeder Technologie aufzulisten. [2] Sie beobachteten eine unglaubwürdig hohe negative Korrelation zwischen zwei Schätzungen, die ihre Probanden machten: dem Nutzwert und der Risikohaftigkeit, die sie den Technologien zuschrieben. Wenn Personen eine positive Einstellung zu einer Technologie hatten, schrieben sie ihr einen großen Nutzen und ein geringes Risiko zu; wenn sie eine Technologie nicht mochten, überwogen die Nachteile, die ihnen dazu einfielen, bei Weitem ihre Vorteile. Weil die Technologien klar als positiv oder negativ eingestuft wurden, waren keine schmerzlichen Kompromisse erforderlich. Risiko-Nutzen-Schätzungen entsprachen sich sogar noch mehr, wenn die Probanden Risiken und Nutzen unter Zeitdruck beurteilten. Bemerkenswerterweise reagierten Mitglieder der British Toxicology Society ganz ähnlich: Substanzen oder Technologien, die sie für riskant hielten, stuften sie als wenig nützlich ein und umgekehrt.[3] Affektkonsistenz ist ein zentrales Element dessen, was ich assoziative Kohärenz genannt habe.

Der beste Teil des Experiments kam als Nächstes. Nach Abschluss der ersten Befragung lasen die Probanden kurze Passagen mit Argumenten zugunsten verschiedener Technologien. Einigen wurden Argumente vorgelegt, die sich auf die vielfältigen Vorteile einer Technologie bezogen, andere bekamen Argumente, die die geringen Risiken betonten. Diese Nachrichten änderten erfolgreich die emotionale Attraktivität der Technologien. Erstaunlicherweise veränderte sich bei Menschen, die eine Nachricht erhalten hatten, in der die Vorteile einer Technologie gepriesen wurden, auch ihre Einstellung zu deren Risiken. Obgleich sie keine sachdienlichen Informationen erhalten hatten, nahmen sie die Technologie, die ihnen jetzt besser gefiel als zuvor, auch als weniger riskant wahr. In ähnlicher Weise entwickelten Versuchspersonen, denen lediglich mitgeteilt wurde, die Risiken einer Technologie seien gering, eine positivere Einschätzung ihres Nutzens. Die Schlussfolgerung, die sich daraus ergibt, ist klar: »Der emotionale Schwanz wedelt mit dem rationalen Hund«, wie der Psychologe Jonathan Haidt in anderem Zusammenhang schrieb.[4] Die Affektheuristik vereinfacht unser Leben, indem sie eine Welt erschafft, die viel geordneter ist als die Wirklichkeit. In der imaginären Welt, in der wir

leben, haben gute Technologien nur geringe Kosten, schlechte Technologien haben keinen Nutzen, und alle Entscheidungen sind leicht. In der realen Welt dagegen müssen wir Nutzen und Kosten oftmals schmerzlich gegeneinander abwägen.

## Die Öffentlichkeit und die Experten

Paul Slovic weiß vermutlich mehr über die Eigentümlichkeiten unserer Risikobeurteilungen als irgendjemand sonst. Seine Arbeiten vermitteln ein Bild von Herr und Frau Bürger, das alles andere als schmeichelhaft ist: gelenkt von Emotionen statt von Vernunft, leicht beeinflussbar durch triviale Details und in unzureichendem Maße empfänglich für Unterschiede zwischen niedrigen und vernachlässigbar niedrigen Wahrscheinlichkeiten. Slovic hat auch Experten erforscht, die wesentlich besser mit Zahlen und Mengen umgehen können. Experten zeigen viele der gleichen Verzerrungen wie die Normalsterblichen in abgeschwächter Form, aber ihre Risikobewertungen und -präferenzen unterscheiden sich oftmals von denen anderer Menschen.

Unterschiede zwischen Experten und der allgemeinen Bevölkerung werden zum Teil mit Verzerrungen bei Urteilen von Laien erklärt, aber Slovic lenkt die Aufmerksamkeit auf Situationen, in denen sich in den Unterschieden ein echter Wertekonflikt widerspiegelt. Er weist darauf hin, dass Experten Risiken oftmals anhand der Anzahl der verlorenen Menschenleben (oder Lebensjahre) messen, während die allgemeine Öffentlichkeit feinere Unterschiede macht, zum Beispiel zwischen »guten Todesfällen« und »schlechten Todesfällen« oder zwischen zufälligen Unfalltoten und Todesfällen, die sich bei Freizeitaktivitäten wie Skifahren ereignen. Diese legitimen Unterscheidungen werden in der Statistik, die bloß Fälle zählt, oftmals ignoriert. Solche Beobachtungen veranlassten Slovic zu der Aussage, Laien hätten ein umfassenderes Risikoverständnis als Experten. Folglich widersetzt er sich entschieden jeglicher »Expertokratie« und der Auffassung, dass Expertenmeinungen kritiklos akzeptiert werden sollten, wenn sie den Meinungen und Wünschen anderer Bürger zuwiderlaufen. Wenn Experten und die Öffentlichkeit unterschiedliche Prioritäten setzten, dann müsse »jede Seite die Erkenntnisse und die Intelligenz der anderen respektieren«, so Slovic.

In seinem Wunsch, die Risikopolitik der alleinigen Kontrolle der Experten zu entwinden, hat Slovic die Grundlage ihrer Expertise (ihres Expertenwissens) infrage gestellt: die Annahme, Risiken seien objektiv.

> »Risiken« existieren nicht »da draußen«, unabhängig von unserem Intellekt und unserer Kultur, und warten darauf, gemessen zu werden. Menschen haben den Begriff des »Risikos« erfunden, der ihnen helfen soll, die Gefahren und Ungewissheiten des Lebens zu verstehen und zu meistern. Obwohl diese Gefahren real sind, gibt es kein »reales Risiko« oder »objektives Risiko«.[5]

Um seine Behauptung zu veranschaulichen, listet Slovic neun mögliche Definitionen des mit der Freisetzung von Giftstoffen in die Luft verbundenen Sterblichkeitsrisikos auf, angefangen von »Todesfälle pro eine Million Menschen« bis zu »Todesfälle pro Millionen Dollar an hergestellten Produkten«. Er will damit zeigen, dass die Risikobewertung von der Wahl eines Maßstabs abhängt – mit der offensichtlichen Möglichkeit, dass die Wahl von einer Präferenz für ein bestimmtes Ergebnis geleitet wurde. Er folgert daraus, dass »Risiken zu definieren Macht auszuüben bedeutet«. Hätten Sie gedacht, dass experimentelle Studien über die psychologischen Grundlagen von Urteilsprozessen einen zu solchen heiklen politischen Fragen führen können? Doch in der Politik geht es letztlich um die Wünsche und das Wohl der Menschen. In jede politische Frage fließen bestimmte Annahmen über die menschliche Natur ein, insbesondere über die Entscheidungen, die Menschen treffen, und die Folgen ihrer Entscheidungen für sich selbst und für die Gesellschaft.

Ein anderer Wissenschaftler und Freund, den ich sehr bewundere, Cass Sunstein, lehnt Slovics Geringschätzung des Experten entschieden ab und verteidigt die Rolle von Experten als ein Bollwerk gegen »populistische« Auswüchse. Sunstein ist einer der bedeutendsten US-amerikanischen Rechtswissenschaftler, und er teilt mit anderen führenden Vertretern seines Fachs die Eigenschaft intellektueller Furchtlosigkeit. Er weiß, dass er sich in jedes Fachgebiet schnell und gründlich einarbeiten kann, und er kennt sich in vielen Bereichen wie etwa der Psychologie von Urteils- und Entscheidungsprozessen, in Fragen der Regulierung und der Risikopolitik hervorragend aus. Seiner Auffassung nach zeichnet sich das bestehende Regulierungssystem in den Vereinigten Staaten durch eine äußerst

mangelhafte Prioritätensetzung aus, in der sich eher die Reaktion auf öffentlichen Druck als eine gründliche objektive Analyse widerspiegelt. Er geht von dem Standpunkt aus, dass Risikoregulierung und staatliche Eingriffe zur Risikominderung einer rationalen Kosten-Nutzen-Abwägung folgen sollten und dass die natürlichen Einheiten für diese Analyse die Anzahl der geretteten Menschenleben (oder auch die Anzahl der geretteten Lebensjahre, was der Rettung junger Menschen mehr Gewicht verleiht) und die Kosten (in Dollar) für die Wirtschaft sein sollten. Eine schlechte Regulierung bedeutet Vergeudung von Menschenleben und Geld, was sich beides objektiv messen lässt. Slovics Argument, dass Risiken und ihre Messung subjektiv seien, hat Sunstein nicht überzeugt. Viele Aspekte der Risikobewertung seien fragwürdig, aber er glaubt fest an die Objektivität, die durch Wissenschaft, Expertenwissen und gründliche Diskussion zur Konsensfindung erreicht werden könne.

Sunstein gelangte zu der Überzeugung, dass verzerrte Reaktionen auf Risiken eine wichtige Ursache willkürlicher und nicht sachgerechter Prioritäten in der Politik sind. Gesetzgeber und Regulierungsbehörden mögen übermäßig empfänglich sein für irrationale Anliegen der Bürger, sowohl wegen der politischen Sensibilität als auch deshalb, weil sie den gleichen kognitiven Verzerrungen unterliegen wie die Bürger.

Sunstein und ein Mitarbeiter, der Jurist Timur Kuran, erfanden einen Namen für den Mechanismus, durch den Verzerrungen in die Politik einfließen: die »Verfügbarkeitskaskade«.[6] Sie sagen, im sozialen Kontext seien »alle Heuristiken gleich, aber die Verfügbarkeitsheuristik ist gleicher«. Sie haben dabei einen erweiterten Heuristikbegriff im Sinn, bei dem die Verfügbarkeit eine Heuristik für Urteile bereitstellt, die sich nicht auf Häufigkeiten beziehen. Insbesondere wird die Bedeutung einer Idee oftmals nach der Flüssigkeit (und der emotionalen Färbung), mit der diese Idee aus dem Gedächtnis abgerufen wird, beurteilt.

Eine Verfügbarkeitskaskade ist eine sich selbst tragende Kette von Ereignissen, die vielleicht mit Medienberichten über ein relativ unbedeutendes Ereignis beginnt und zu öffentlicher Panik und massiven staatlichen Maßnahmen führt. Manchmal zieht ein Medienbericht die Aufmerksamkeit eines Teils der Öffentlichkeit auf sich, die dadurch aufgerüttelt und beunruhigt wird. Diese emotionale Reaktion wird selbst zu einer Geschichte, die ihrerseits weitere Berichte in den Medien auslöst, was noch größere Besorgnis und Engagement hervorruft. Dieser Kreislauf wird

manchmal gezielt von »Verfügbarkeitsunternehmern« beschleunigt; das sind Einzelpersonen oder Organisationen, die daran arbeiten, einen beständigen Fluss beunruhigender Nachrichten aufrechtzuerhalten. Die Gefahr wird in dem Maße überzeichnet, wie die Medien um reißerische Schlagzeilen konkurrieren. Wissenschaftler und andere, die versuchen, die wachsende Angst und Abscheu zu dämpfen, finden wenig Beachtung, und die meiste davon ist feindseliger Natur: Jeder, der behauptet, die Gefahr werde übertrieben, wird der Beteiligung an einer »ruchlosen Vertuschung« verdächtigt. Die Sache gewinnt politische Bedeutung, weil sie die Öffentlichkeit umtreibt, und die Reaktion des politischen Systems wird von der Intensität der öffentlichen Stimmung determiniert. Die Verfügbarkeitskaskade hat jetzt die Prioritäten neu festgesetzt. Weitere Risiken und andere das Gemeinwohl fördernde Verwendungsweisen von Ressourcen sind in den Hintergrund getreten.

Kuran und Sunstein konzentrierten sich auf zwei Beispiele, die noch immer umstritten sind: die Love-Canal-Affäre und den sogenannten Alar-Zwischenfall. Bei der Love-Canal-Affäre im Jahr 1979 wurde durch ungewöhnlich starke Niederschläge vergrabener Giftmüll freilegt, mit der Folge, dass das Trinkwasser nicht nur faulig stank, sondern auch in gesundheitsgefährdender Weise mit Schadstoffen belastet war. Die Bewohner der Gemeinde waren wütend und besorgt, und einer von ihnen, Lois Gibbs, bemühte sich besonders aktiv darum, das Interesse an dem Problem aufrechtzuerhalten. Die Verfügbarkeitskaskade entfaltete sich drehbuchmäßig. Auf dem Höhepunkt des Skandals berichteten die Medien täglich über Love Canal. Wissenschaftler, die behaupteten, die Gefahren würden überzeichnet, wurden ignoriert oder niedergebrüllt, ABC News strahlte eine Sendung mit dem Titel *Todesboden* aus, und leere, babygroße Särge wurden vor dem Parlament zur Schau gestellt. Viele Bewohner wurden auf Staatskosten umgesiedelt, und die Überwachung und sichere Entsorgung von Giftmüll wurde zur wichtigsten Umweltfrage der 1980er-Jahre. Das Gesetz, das die Sanierung von Giftmülllagerstätten anordnete, genannt CERCLA, führte zu einem Superfonds und gilt als ein Meilenstein der Umweltgesetzgebung. Es war kostspielig, und einige haben behauptet, die gleiche Summe hätte viel mehr Menschenleben retten können, wenn sie für andere Zwecke ausgegeben worden wäre. Die Meinungen darüber, was eigentlich genau in Love Canal passiert ist, gehen noch immer weit auseinander, und Behauptungen über Gesundheitsschäden scheinen sich

nicht bewahrheitet zu haben. Kuran und Sunstein schildern die Love-Canal-Affäre fast als ein Pseudoereignis, während auf der anderen Seite der Debatte Umweltschützer noch immer von der »Love-Canal-Katastrophe« sprechen.

Auch bei dem zweiten Beispiel, das Kuran und Sunstein benutzten, um ihr Konzept einer Verfügbarkeitskaskade zu veranschaulichen, gehen die Meinungen weit auseinander: dem Alar-Zwischenfall von 1989, der von Kritikern der Umweltschutzbewegung auch »Alar-Hysterie« genannt wird. Alar ist eine Chemikalie, mit der Äpfel besprüht wurden, um ihr Wachstum zu regulieren und ihr optisches Erscheinungsbild zu verbessern. Die Hysterie begann mit Presseberichten, wonach die Chemikalie bei Ratten und Mäusen, die sie in gigantischen Mengen aufgenommen hatten, bösartige Tumoren verursache. Die Berichte verängstigten verständlicherweise die Öffentlichkeit, und diese Befürchtungen lösten weitere Medienberichte aus, die der Grundmechanismus einer Verfügbarkeitskaskade sind. Das Thema beherrschte die Nachrichten und produzierte dramatische Medienereignisse, wie etwa die Aussage der Schauspielerin Meryl Streep vor einem Kongressausschuss. Die Apfelwirtschaft erlitt große Verluste, da Äpfel und Apfelprodukte zu Objekten der Furcht wurden. Kuran und Sunstein zitieren einen Bürger, der sich erkundigte, »ob es sicherer ist, Apfelsaft in den Ausguss zu gießen oder ihn zur Sondermülldeponie zu bringen«. Der Hersteller nahm das Produkt vom Markt, und die FDA – die US-Behörde für Lebensmittelsicherheit – verbot die Substanz. Spätere Forschungen bestätigten, dass der Stoff möglicherweise das Krebsrisiko ganz geringfügig erhöht, aber der Alar-Zwischenfall war zweifellos eine gewaltige Überreaktion auf ein geringfügiges Problem. Insgesamt wirkte sich der Vorfall wohl eher negativ auf die Volksgesundheit aus, weil weniger gesundheitsfördernde Äpfel verzehrt wurden.

Die Alar-Geschichte verdeutlicht eine grundlegende Begrenzung in der Fähigkeit unseres Geistes, kleine Risiken richtig zu bewerten: Entweder wir ignorieren sie ganz und gar, oder wir überschätzen sie maßlos – dazwischen gibt es nichts.[7] Jeder Vater oder jede Mutter, der/die schon einmal aufgeblieben ist, um zu warten, bis die heranwachsende Tochter, die schon längst von einer Party zurück sein sollte, heimkommt, dürfte dieses Gefühl kennen. Vielleicht wissen Sie, dass Sie eigentlich (so gut wie) keinen Grund zur Sorge haben, aber Sie können nichts dagegen tun, dass sich Ihnen Bilder

von einem schlimmen Unglück aufdrängen. Slovic hat behauptet, das Ausmaß der Besorgtheit sei nicht hinlänglich empfindlich für die Schadenswahrscheinlichkeit; man malt sich den Zähler aus – die tragische Geschichte, die man gerade in den Nachrichten gesehen hat – und denkt nicht über den Nenner nach. Sunstein hat den Ausdruck »Wahrscheinlichkeitsvernachlässigung« geprägt, um dieses Muster zu beschreiben. Die Kombination der Wahrscheinlichkeitsvernachlässigung mit den sozialen Mechanismen von Verfügbarkeitskaskaden führt zwangsläufig zu einer erheblichen Überbewertung unwahrscheinlicher Bedrohungen, manchmal mit weitreichenden Folgen.

In der heutigen Welt sind Terroristen die bedeutendsten Praktiker der Kunst, Verfügbarkeitskaskaden auszulösen. Abgesehen von einigen wenigen schrecklichen Ausnahmen, wie den Anschlägen vom 11. September 2001, ist die Anzahl der Opfer von Terroranschlägen sehr gering im Vergleich zu anderen Todesursachen. Selbst in Ländern, die Ziele intensiver Terrorkampagnen waren, wie etwa Israel, erreichte die wöchentliche Zahl der Todesopfer praktisch niemals auch nur annähernd die Zahl der Unfalltoten. Der Unterschied liegt in der Verfügbarkeit der beiden Risiken, der Leichtigkeit und Häufigkeit, mit denen sie uns einfallen. Grauenvolle Bilder, die in den Medien endlos wiederholt werden, nehmen uns alle schwer mit. Wie ich aus Erfahrung weiß, ist es schwer, sich durch vernünftiges Argumentieren selbst in einen Zustand völliger Ruhe zu versetzen. Terrorismus spricht direkt System 1 an.

Wo stehe ich in der Debatte zwischen meinen Freunden? Verfügbarkeitskaskaden sind real, und sie verzerren zweifellos die Prioritäten bei der Zuteilung öffentlicher Ressourcen. Cass Sunstein würde nach Mechanismen suchen, die Entscheidungsträger vor öffentlichem Druck schützen, und unparteiische Experten, die einen Überblick über alle Risiken und die zu ihrer Reduzierung verfügbaren Ressourcen haben, über die Zuteilung von Ressourcen entscheiden lassen. Paul Slovic vertraut Experten viel weniger und der Öffentlichkeit etwas mehr, als es Sunstein tut, und er weist darauf hin, dass die Isolierung der Experten von den Emotionen der Bürger eine Politik erzeugt, die die Bevölkerung ablehnen wird – eine unhaltbare Situation in einer Demokratie. Beide Standpunkte sind durch und durch vernünftig, und ich schließe mich beiden an.

Ich teile Sunsteins Unbehagen über den Einfluss irrationaler Ängste und Verfügbarkeitskaskaden auf den politischen Umgang mit Risiken. Aber ich

teile auch Slovics Ansicht, dass weitverbreitete Ängste, selbst wenn sie unbegründet sind, von Politikern nicht ignoriert werden sollten. Rational oder nicht, Ängste sind unangenehm und kraftraubend, und Politiker müssen bestrebt sein, die Bevölkerung nicht nur gegen reale Gefahren, sondern auch vor Ängsten zu schützen.

Slovic betont zu Recht, dass es die Bürger nicht hinnehmen werden, wenn Entscheidungen von nicht gewählten und nicht rechenschaftspflichtigen Experten getroffen werden. Außerdem können Verfügbarkeitskaskaden langfristig vorteilhaft sein, indem sie die Aufmerksamkeit auf Risikoklassen lenken und die Gesamthöhe des Budgets zur Risikominderung erhöhen. Der Love-Canal-Vorfall mag dazu geführt haben, dass unverhältnismäßig viele Ressourcen für die Entsorgung von Giftmüll aufgewendet wurden, aber er hatte auch den allgemeineren Effekt, dass ökologische Belange einen höheren Stellenwert erhielten. Demokratische Entscheidungsfindung ist zwangsläufig eine vertrackte Angelegenheit, zum Teil deshalb, weil die Verfügbarkeits- und die Affektheuristik, die die Überzeugungen und Einstellungen der Bürger prägen, zwangsläufig verzerrt sind, auch wenn sie im Allgemeinen in die richtige Richtung weisen. Psychologische Erkenntnisse sollten bei der Konzipierung von Risikopolitiken, die Expertise mit Emotionen und Intuitionen der Öffentlichkeit verknüpfen, berücksichtigt werden.

## Zum Thema »Verfügbarkeitskaskaden«

»Sie schwärmt von einer Innovation, die einen großen Nutzen und keine Kosten haben soll. Ich vermute, das liegt an der Affektheuristik.«

»Dies ist eine Verfügbarkeitskaskade: Ein Nichtereignis, das von den Medien und der Öffentlichkeit aufgeblasen wurde, bis

es unsere Fernsehbildschirme füllt und zum alleinigen Gesprächsthema geworden ist.«

# *14. Was studiert Tom W.?*

Werfen Sie einen Blick auf diese einfache Denkaufgabe:

> Tom W. ist Student an der größten Universität Ihres Bundesstaates. Bitte ordnen Sie die folgenden neun Fachrichtungen in der Reihenfolge der Wahrscheinlichkeit, mit der Tom jetzt jedes dieser Fachgebiete studiert. Verwenden Sie eine Eins für die wahrscheinlichste Fachrichtung und eine Neun für die am wenigsten wahrscheinliche Fachrichtung:
>
> Betriebswirtschaftslehre
> Informatik
> Ingenieurwissenschaft
> Geisteswissenschaften und Pädagogik
> Rechtswissenschaft
> Medizin
> Bibliothekswissenschaft
> Physik- und Biowissenschaften
> Sozialwissenschaften und Sozialarbeit

Die Frage ist leicht, und Sie wussten sofort, dass der prozentuale Anteil der in die einzelnen Fachrichtungen eingeschriebenen Studenten an der Gesamtzahl der Studierenden der Schlüssel zur Lösung ist. Soweit Sie wissen, wurde Tom W. zufällig aus den Studenten der Universität ausgewählt, so wie eine einzelne Murmel aus einer Urne gezogen wird. Um zu entscheiden, ob eine Murmel eher rot oder eher grün ist, müssen wir wissen, wie viele Murmeln jeder Farbe in der Urne sind. Dieser Anteil der Murmeln einer bestimmten Sorte wird »Basisrate« genannt. Entsprechend

ist die Basisrate für Geisteswissenschaften und Pädagogik in diesem Problem der Anteil der Studenten dieser Fachrichtung unter allen Studenten. Wenn keine konkreten Informationen über Tom W. vorhanden sind, hält man sich an die Basisraten und nimmt an, dass er mit höherer Wahrscheinlichkeit Geisteswissenschaften und Pädagogik als Informatik oder Bibliothekswissenschaft studiert, weil es insgesamt mehr Studenten der Geisteswissenschaften und Pädagogik als der anderen beiden Fachrichtungen gibt. Wenn keine weitergehenden Informationen zur Verfügung stehen, bietet es sich an, von den Basisraten auszugehen. Als Nächstes folgt eine Aufgabe, die nichts mit Basisraten zu tun hat.

> Nachfolgend finden Sie eine Persönlichkeitsskizze von Tom W., die ein Psychologe auf der Basis von psychologischen Tests unbekannter Validität (Gültigkeit) in Toms letztem Highschool-Jahr geschrieben hat:
>
> > Tom W. ist hochintelligent, auch wenn es ihm an echter Kreativität mangelt. Er hat ein Bedürfnis nach Ordnung und Klarheit und nach übersichtlichen, strukturierten Systemen, in denen jedes Element seinen geeigneten Platz findet. Sein Schreibstil ist eher fade und mechanisch und wird nur gelegentlich durch etwas abgedroschene Wortspiele und ein kurzes Aufblitzen einer Science-Fiction-artigen Fantasie verlebendigt. Er scheint wenig Gespür für andere Menschen und ein geringes Einfühlungsvermögen zu besitzen, und er ist eher kontaktscheu. Trotz seiner Selbstbezogenheit hat er ein ausgeprägtes moralisches Bewusstsein.
>
> Nehmen Sie jetzt bitte ein Blatt Papier und ordnen Sie die neun unten aufgeführten Fachrichtungen danach, wie ähnlich die Beschreibung von Tom W. dem typischen Studenten in jedem

Fachgebiet ist. Verwenden Sie eine Eins für die stärkste und eine Neun für die geringste Ähnlichkeit.

Sie haben mehr von diesem Kapitel, wenn Sie diese Aufgabe kurz selbst probieren; die Lektüre des Berichts über Tom W. ist notwendig, damit Sie sich ein Urteil über die verschiedenen Fachrichtungen bilden können.

Auch diese Frage ist einfach. Sie verlangt von Ihnen nur, dass Sie das Stereotyp von Studenten verschiedener Studiengänge abrufen oder vielleicht auch konstruieren. Als das Experiment Anfang der 1970er-Jahre zum ersten Mal durchgeführt wurde, sah die Rangfolge, gemittelt über alle Probanden, folgendermaßen aus. Ihre eigene wird vermutlich nicht allzu stark davon abweichen:

1. Informatik
2. Ingenieurwissenschaft
3. Betriebswirtschaftslehre
4. Physik- und Biowissenschaften
5. Bibliothekswissenschaft
6. Rechtswissenschaft
7. Medizin
8. Geisteswissenschaften und Pädagogik
9. Sozialwissenschaften und Sozialarbeit

Vermutlich haben Sie Informatik angesichts der Hinweise auf kauziges Verhalten (»abgedroschene Wortspiele«) zu den passendsten Fächern gezählt. Tatsächlich passt die Beschreibung von Tom W. auf dieses Stereotyp. Ein weiterer Studiengang, der von den meisten weit oben eingeordnet wurde, ist die Ingenieurwissenschaft (»Bedürfnis nach … übersichtlichen, strukturierten Systemen«). Vermutlich glaubten Sie, dass sich das Porträt von Tom W. nicht gut mit Ihrer Vorstellung von Sozialwissenschaften und Sozialarbeit (»wenig Gespür für andere Menschen und ein geringes Einfühlungsvermögen«) zusammenbringen lässt. Berufliche Stereotype scheinen sich in den fast vierzig Jahren, die vergangen sind, seitdem ich diese Beschreibung von Tom W. ausgearbeitet habe, kaum verändert zu haben.

Die Aufgabe, eine Rangordnung der neun Studiengänge zu erstellen, ist komplex und erfordert zweifellos jene Disziplin und sequenzielle

Organisation, zu der nur System 2 in der Lage ist. Doch die in die Beschreibung eingestreuten Hinweise (»abgedroschene Wortspiele« und andere) sollten eine Assoziation mit einem Stereotyp aktivieren, was eine automatische Operation von System 1 ist.

Die Anweisungen für diese Ähnlichkeitsaufgabe verlangten einen Vergleich der Beschreibung von Tom W. mit den Stereotypen verschiedener Fachrichtungen. Für die Lösung dieser Aufgabe spielt es keine Rolle, ob die Beschreibung zutreffend ist – ob es sich um ein wahres Porträt von Tom W. handelt oder nicht. Und das Gleiche gilt auch für unser Wissen über die Basisraten der verschiedenen Studiengänge. Die Ähnlichkeit einer Person mit dem Stereotyp einer Gruppe wird nicht von der Größe dieser Gruppe beeinflusst. Tatsächlich könnten Sie die Beschreibung von Tom auch dann mit dem Stereotyp von Studenten der Bibliothekswissenschaft vergleichen, wenn es an der Universität gar keinen solchen Studiengang gibt.

Wenn Sie die Beschreibung von Tom W. noch einmal durchlesen, werden Sie feststellen, dass sie gut zu Stereotypen von kleinen Gruppen von Studenten (Informatikern, Bibliothekswissenschaftlern, Ingenieuren) passt, während sie mit denen der größten Gruppen (Geisteswissenschaften und Pädagogik, Sozialwissenschaften und Sozialarbeit) viel schlechter zusammenpasst. Tatsächlich haben die Probanden die beiden größten Studiengänge fast immer sehr niedrig eingestuft. Tom W. wurde gezielt als eine »Anti-Basisraten«-Figur konzipiert, die gut mit den Stereotypen kleiner Studiengänge, aber schlecht mit denen der größten Studiengänge zusammenpasst.

## Vorhersage durch Repräsentativität

Die dritte Aufgabe in der Reihe wurde Psychologie-Studenten gestellt, und sie ist die entscheidende: Ordnen Sie die Studiengänge nach der Wahrscheinlichkeit, mit der Tom W. heute das jeweilige Fach studiert. Die Mitglieder dieser Vorhersagegruppe kannten die relevanten statistischen Daten: Sie wussten die Basisraten der verschiedenen Studiengänge, und sie wussten, dass die Quelle von Tom W.s Beschreibung nicht in hohem Maße vertrauenswürdig ist. Allerdings erwarteten wir, dass sie sich ausschließlich auf die Ähnlichkeit der Beschreibung mit den Stereotypen konzentrieren – auf die »Repräsentativität«, wie wir dies nannten – und sowohl die Basisraten als auch die Zweifel an der Richtigkeit der Beschreibung ausblenden würden. Anschließend würden sie den Studiengang mit wenigen Studierenden – Informatik – als am wahrscheinlichsten beurteilen, weil dieses Ergebnis den höchsten Repräsentativitätswert erhält.

Amos und ich arbeiteten hart während des Jahres, das wir in Eugene verbrachten, und ich blieb manchmal die ganze Nacht in meinem Büro. Eine meiner Aufgaben in diesen Nächten bestand darin, mir eine Beschreibung auszudenken, die die Repräsentativität und die Basisraten gegeneinander ausspielen sollte. Tom W. war das Produkt meiner Anstrengungen, und ich stellte die Beschreibung in den frühen Morgenstunden fertig. Die erste Person, die an jenem Morgen zur Arbeit erschien, war unser Kollege und Freund Robyn Dawes, der nicht nur ein scharfsinniger Statistiker war, sondern auch große Zweifel an der Aussagekraft intuitiver Urteile hatte. Wenn jemand die Bedeutung der Basisrate kennen musste, dann war es Robyn. Ich rief ihn her, zeigte ihm die Frage, die ich gerade getippt hatte, und bat ihn, Tom W.s Beruf zu erraten. Ich erinnere mich noch immer an sein listiges Lächeln, als er zaghaft antwortete: »Informatiker ?« Dies war ein glücklicher Augenblick – selbst der Experte war darauf hereingefallen. Natürlich erkannte Robyn sofort seinen Fehler, als ich »Basisrate« erwähnte, aber er hatte nicht von sich aus daran gedacht. Obwohl er bestens über die Rolle von Basisraten bei Vorhersagen Bescheid wusste, vernachlässigte er sie, als ihm die Beschreibung der Persönlichkeit eines Individuums vorgelegt wurde. Wie

erwartet, ersetzte er die Wahrscheinlichkeit, nach der er gefragt worden war, durch ein Repräsentativitätsurteil.

Anschließend sammelten Amos und ich die Antworten auf dieselbe Frage von 114 Psychologie-Studenten an drei großen Universitäten, die alle mehrere Statistikkurse besucht hatten. Sie enttäuschten uns nicht. Ihre Einstufung der neun Studiengänge nach der Wahrscheinlichkeit unterschied sich nicht von der Beurteilung gemäß der Stereotyp-Ähnlichkeit. In diesem Fall war die Ersetzung perfekt: Es gab keinerlei Anhaltspunkte dafür, dass die Probanden etwas anderes machten, als die Repräsentativität zu beurteilen. Die Frage nach der Wahrscheinlichkeit *(likelihood)* war schwierig, die Frage nach der Ähnlichkeit war leichter, und so wurde sie stattdessen beantwortet. Dies ist ein gravierender Fehler, weil Ähnlichkeits- und Wahrscheinlichkeitsurteile nicht den gleichen logischen Regeln unterliegen. Es ist völlig in Ordnung, wenn Basisraten und die Möglichkeit, dass die Beschreibung unzutreffend ist, keinen Einfluss auf Ähnlichkeitsurteile haben, aber jeder, der bei Wahrscheinlichkeitseinschätzungen Basisraten und die Qualität der Informationen außer Acht lässt, wird zweifellos Fehler machen.

Das Konzept von der »Wahrscheinlichkeit, dass Tom W. Informatik studiert« ist kein einfaches. Logiker und Statistiker sind sich über seine Bedeutung uneins, und einige würden sogar sagen, dass es gar keine Bedeutung hat. Für viele Experten ist es ein Maß für den Grad des subjektiven Überzeugtseins. Es gibt einige Ereignisse, von denen man fest überzeugt ist, zum Beispiel dass die Sonne heute Morgen aufgegangen ist, und andere, die man für unmöglich hält, wie etwa, dass der Pazifik auf einen Schlag völlig zufriert. Dann gibt es viele Ereignisse, wie etwa die Vermutung, dass Ihr Wohnungsnachbar ein Informatiker ist, denen man einen mittleren Grad des Überzeugtseins zuweist – was Ihrem subjektiven Wahrscheinlichkeitswert dieses Ereignisses entspricht.

Logiker und Statistiker haben konkurrierende Definitionen der Wahrscheinlichkeit entwickelt, die alle sehr präzise sind. Für Laien jedoch ist Wahrscheinlichkeit ein vager Begriff, der mit Ungewissheit, Neigung *( propensity),* Plausibilität und Überraschung zusammenhängt. Vagheit ist nicht auf dieses Konzept beschränkt, und sie ist auch nicht besonders störend. Wir wissen mehr oder minder, was wir meinen, wenn wir ein Wort wie »Demokratie« oder »Schönheit« verwenden, und die Menschen, mit denen wir sprechen, verstehen mehr oder weniger, was wir sagen wollen. In

all den Jahren, in denen ich Fragen über die Wahrscheinlichkeit von Ereignissen stellte, hat niemand jemals eine Hand gehoben, um mich zu fragen: »Sir, was verstehen Sie unter Wahrscheinlichkeit ?«, wie sie es getan hätten, wenn ich sie gebeten hätte, ein unbekanntes Konzept wie »Globabilität« zu beurteilen. Alle verhielten sich so, als wüssten sie die Antwort auf meine Fragen, auch wenn uns allen klar ist, dass es unfair wäre, sie zu bitten, die Bedeutung des Wortes zu erklären.

Menschen, die die Wahrscheinlichkeit von Ereignissen beurteilen sollen, sind nicht überfordert, weil sie nicht versuchen, die Wahrscheinlichkeit in dem Sinne zu beurteilen, in dem Statistiker und Philosophen den Begriff verwenden. Eine Frage nach der Wahrscheinlichkeit aktiviert eine mentale Schrotflinte, die Antworten auf leichtere Fragen hervorruft. Eine der leichten Antworten ist eine automatische Beurteilung der Repräsentativität – ein übliches Verfahren beim Sprachverstehen. Die (falsche) Aussage »Elvis Presleys Eltern wollten, dass er Zahnarzt wird« wirkt leicht komisch, weil die Diskrepanz zwischen dem Image von Presley und dem Stereotyp eines Zahnarztes automatisch erkannt wird. System 1 erzeugt einen Eindruck von Ähnlichkeit, ohne dies zu beabsichtigen. Die Repräsentativitätsheuristik kommt ins Spiel, wenn jemand »Sie wird die Wahlen gewinnen; man sieht ihr an, dass sie eine Gewinnerin ist« oder »Er wird es als Akademiker nicht weit bringen; er hat zu viele Tätowierungen« sagt. Wir stützen uns auf Repräsentativität, wenn wir die potenzielle Führungsstärke eines Bewerbers um ein Amt anhand der Form seines Kinns oder der Eindringlichkeit seiner Reden beurteilen.

Obwohl es weit verbreitet ist, ist es statistisch nicht optimal, Vorhersagen auf der Basis der Repräsentativität zu machen. Michael Lewis’ Bestseller *Moneyball* ist eine Geschichte über die Ineffizienz dieser Vorhersagemethode. Professionelle Baseball-Scouts sagen die Leistungsfähigkeit potenzieller Spieler traditionell zum Teil auf der Basis ihres Körperbaus und ihres Aussehens vorher. Der Held von Lewis’ Buch ist Billy Beane, der Trainer der Oakland A ’s, der die unpopuläre Entscheidung trifft, sich über die Empfehlungen seiner Scouts hinwegzusetzen und Spieler nach Maßgabe der statistischen Daten ihrer bisherigen Leistungen auszuwählen. Die Spieler, für die er sich entschied, waren billig, weil andere Teams sie wegen ihres Aussehens verschmäht hatten. Die Mannschaft verzeichnete schon bald großartige Erfolge zu niedrigen Kosten.

## Die Sünden der Repräsentativität

Es hat gewichtige Vorteile, wenn man Wahrscheinlichkeiten auf der Basis der Repräsentativität beurteilt: Die intuitiven Eindrücke, die auf diesem Wege entstehen, sind oft – in aller Regel – zutreffender, als es Zufallsschätzungen wären.

– Meistens sind Menschen, die sich freundlich verhalten, tatsächlich freundlich.
– Ein Profisportler, der sehr groß und schlank ist, ist mit viel höherer Wahrscheinlichkeit ein Basketball- als ein Footballspieler.
– Menschen mit einem Doktortitel abonnieren eher die *New York Times* als Menschen, die nur einen Highschool-Abschluss haben.
– Junge Männer fahren eher aggressiv als ältere Frauen.

In all diesen und in vielen anderen Fällen ist an den Stereotypen, die Repräsentativitätsurteilen zugrunde liegen, etwas dran, und Vorhersagen auf der Grundlage dieser Heuristik können zutreffend sein. In anderen Situationen sind diese Stereotype falsch, und die Repräsentativitätsheuristik ist irreführend, wenn sie Menschen dazu veranlasst, Informationen über Basisraten zu vernachlässigen, die in eine andere Richtung weisen. Selbst wenn die Heuristik eine begrenzte Gültigkeit besitzt, führt es zu schweren Verstößen gegen die statistische Logik, wenn man sich ausschließlich darauf verlässt.

Eine Sünde der Repräsentativität ist die Bereitschaft, die Häufigkeit unwahrscheinlicher Ereignisse (mit niedriger Basisrate) zu überschätzen. Hier ist ein Beispiel: Sie sehen in der New Yorker U-Bahn eine Person, die die *New York Times* liest. Welche der folgenden Aussagen ist eine bessere Wette über die lesende Unbekannte?

Sie hat einen Doktortitel.
Sie hat keinen College-Abschluss.

Die Repräsentativität würde Ihnen nahelegen, auf den Doktortitel zu wetten, aber das ist nicht unbedingt klug. Sie sollten ernsthaft die zweite Alternative in Erwägung ziehen, weil viel mehr Nichtakademiker als Doktoren die New Yorker U-Bahn benutzen. Und wenn Sie raten sollen, ob eine Frau, die als »schüchterne Poesie-Liebhaberin« beschrieben wird, chinesische Literatur oder Betriebswirtschaftslehre studiert, sollten Sie sich für Letzteres entscheiden.[1] Selbst wenn jede Studentin chinesischer Literatur schüchtern ist und Gedichte liebt, ist es fast sicher, dass es in der viel größeren Population von BWL-Studentinnen mehr scheue Poesie-Liebhaberinnen gibt.

Menschen ohne Statistikkenntnisse sind durchaus in der Lage, sich unter bestimmten Bedingungen bei Vorhersagen auf Basisraten zu stützen. In der ersten Fassung des Tom-W.-Problems, in der keine detaillierten Informationen über ihn enthalten sind, liegt es für jeden auf der Hand, dass die Wahrscheinlichkeit, dass Tom W. eine bestimmte Fachrichtung studiert, einfach der relativen Häufigkeit aller in dieses Fach eingeschriebenen Studenten entspricht. Doch sobald Tom W.s Persönlichkeit beschrieben wird, schwindet das Interesse an den Basisraten augenscheinlich.

Amos und ich glaubten auf der Basis unserer anfänglichen Daten zunächst, dass Informationen über die Basisrate *immer* vernachlässigt werden, wenn Informationen über den konkreten Fall verfügbar sind, aber diese Schlussfolgerung ging zu weit. Psychologen haben viele Experimente durchgeführt, in denen Informationen über Basisraten ausdrücklich als Teil des Problems bereitgestellt werden, und viele Teilnehmer ließen sich von diesen Basisraten beeinflussen, obwohl die Informationen über den Einzelfall immer stärker gewichtet wurden als bloße statistische Angaben.[2] Norbert Schwarz und seine Mitarbeiter haben gezeigt, dass sich die Nutzung der Basisraten-Informationen verbesserte, wenn die Probanden angewiesen wurden, »wie ein Statistiker zu denken«, während die Anweisung, »wie ein Kliniker zu denken«, die gegenteilige Wirkung hatte.[3]

Ein Experiment, das vor ein paar Jahren von Harvard-Studenten durchgeführt wurde, erbrachte ein Ergebnis, das mich überraschte: Die verstärkte Aktivierung von System 2 ergab eine deutliche Verbesserung der Vorhersagegenauigkeit beim Tom-W.-Problem. Bei diesem Experiment wurde das alte Problem mit einer modernen Variante der kognitiven Flüssigkeit verknüpft. Die Hälfte der Studenten sollte während der Aufgabe

ihre Wangen aufblasen, während die anderen die Stirn runzeln sollte.[4] Wie schon gesehen, steigert Stirnrunzeln im Allgemeinen die Vigilanz von System 2 und vermindert sowohl übermäßiges Vertrauen als auch den Rückgriff auf Intuitionen. Die Studenten, die ihre Wangen aufbliesen (ein emotional neutraler Ausdruck), reproduzierten die ursprünglichen Ergebnisse: Sie stützten sich ausschließlich auf Repräsentativität und ignorierten die Basisraten. Die Stirnrunzler dagegen zeigten, wie von den Autoren vorhergesagt, eine gewisse Empfänglichkeit für die Basisraten. Dies ist ein aufschlussreicher Befund.

Bei einem fehlerhaften intuitiven Urteil liegt die Schuld sowohl bei System 1 als auch bei System 2. System 1 schlug die fehlerhafte Intuition vor, und System 2 unterstützte sie und drückte sie in einem Urteil aus. Allerdings gibt es zwei mögliche Gründe für das Versagen von System 2 – Ignoranz oder Faulheit. Einige Menschen ignorieren Basisraten, weil sie sie für irrelevant halten, wenn individuelle Informationen vorliegen. Andere machen den gleichen Fehler, weil sie ihre Aufmerksamkeit nicht auf die Aufgabe fokussieren. Auch wenn das Stirnrunzeln eine Wirkung hat, scheint Faulheit die angemessene Erklärung für die Vernachlässigung der Basisrate zu sein, zumindest bei Harvard-Studenten. Ihr System 2 »weiß«, dass die Basisraten relevant sind, auch wenn sie nicht ausdrücklich erwähnt werden, aber es wendet dieses Wissen nur dann an, wenn es sich für die Aufgabe besonders stark anstrengt.

Die zweite Sünde der Repräsentativität ist Gleichgültigkeit gegenüber der Qualität der Daten. Erinnern wir uns an die Regel von System 1: WYSIATI. In dem Beispiel von Tom W. wird Ihre Assoziationsmaschine durch eine Beschreibung von Tom aktiviert, die diesen zutreffend porträtieren mag oder auch nicht. Die Aussage »Tom W. hat wenig Gespür und wenig Einfühlung in Menschen« genügte vermutlich, um Sie (und die meisten anderen Leser) davon zu überzeugen, dass er wohl kaum ein Student der Sozialwissenschaften oder Sozialarbeit ist. Aber Ihnen wurde ausdrücklich nahegelegt, der Beschreibung nicht zu trauen!

Sie sehen bestimmt grundsätzlich ein, dass wertlose Informationen nicht anders behandelt werden sollten als ein völliges Fehlen von Informationen, aber die WYSIATI-Regel macht es sehr schwer, diesen Grundsatz anzuwenden. Wenn Sie Informationen nicht sofort verwerfen (zum Beispiel dadurch, dass Sie zu dem Schluss kommen, sie von einem Lügner erhalten

zu haben), wird Ihr System 1 die verfügbaren Informationen so behandeln, als wären sie wahr. Wenn Sie Zweifel an der Qualität der Informationen haben, können Sie Folgendes tun: Stellen Sie sicher, dass sich Ihre Wahrscheinlichkeitsurteile in der Nähe der Basisrate bewegen. Erwarten Sie nicht, dass diese Übung in Disziplin leicht ist – sie erfordert erhebliche Anstrengungen bei der Selbstüberwachung und Selbstkontrolle.

Die richtige Antwort für die Tom-W.-Aufgabe lautet: Sie sollten sehr nahe bei Ihren früheren Überzeugungen bleiben, die anfänglich hohen Wahrscheinlichkeiten stark frequentierter Fächer (Geisteswissenschaften und Pädagogik; Sozialwissenschaften und Sozialarbeit) leicht herabsetzen und die niedrigen Wahrscheinlichkeiten schwach frequentierter Studiengänge (Bibliothekswissenschaft, Informatik) leicht erhöhen. Sie sind nicht genau dort, wo Sie wären, wenn Sie überhaupt nichts über Tom W. wüssten, aber die wenigen Informationen, die Sie haben, sind nicht vertrauenswürdig, sodass sich Ihre Schätzungen weitgehend an die Basisraten anlehnen sollten.

## Wie man die Intuition diszipliniert

Die von Ihnen geschätzte Wahrscheinlichkeit, dass es morgen regnen wird, ist Ihr subjektiver Grad der Überzeugung, aber Sie sollten nicht alles glauben, was Ihnen einfällt. Um nützlich zu sein, sollten Ihre Überzeugungen der statistischen Logik der Wahrscheinlichkeit genügen. Wenn Sie zum Beispiel glauben, dass es mit einer Wahrscheinlichkeit von 40 Prozent morgen regnen wird, müssen Sie auch glauben, dass es mit einer Wahrscheinlichkeit von 60 Prozent morgen nicht regnen wird, und Sie dürfen nicht glauben, dass es mit einer Wahrscheinlichkeit von 50 Prozent morgen früh regnen wird. Und wenn Sie glauben, dass Kandidat X mit einer Wahrscheinlichkeit von 30 Prozent zum Präsidenten gewählt wird und eine 80-prozentige Chance hat, wiedergewählt zu werden, dann müssen Sie glauben, dass die Wahrscheinlichkeit, dass er zweimal hintereinander gewählt wird, 24 Prozent beträgt.

Die relevanten »Regeln« für Fälle wie das Tom-W.-Problem stellt die Bayessche Statistik bereit. Diese einflussreiche moderne Statistiktheorie ist nach einem englischen Geistlichen des 18. Jahrhunderts benannt, Reverend Thomas Bayes, der den ersten bedeutenden Beitrag zur Lösung eines großen Problems lieferte: der Logik, wie Menschen im Lichte neuer Erkenntnisse ihre Einstellung ändern sollten. Die Bayessche Regel gibt genau an, wie bestehende Überzeugungen (in den Beispielen dieses Kapitels waren dies die Basisraten) mit der Aussagekraft der Erkenntnisse – dem Ausmaß, in dem sie die Hypothese gegenüber der Alternative begünstigen – kombiniert werden sollten.[5] Wenn Sie zum Beispiel glauben, dass 3 Prozent der Studenten Informatik studieren (die Basisrate), und Sie auch glauben, dass die Beschreibung von Tom W. mit vierfach höherer Wahrscheinlichkeit auf einen Studenten dieser Fachrichtung als auf Studenten anderer Studiengänge zutrifft, dann besagt die Bayessche Regel, dass die Wahrscheinlichkeit, dass Tom W. Informatiker ist, Ihrer Überzeugung nach jetzt 11 Prozent betragen muss. Wenn die Basisrate 80 Prozent betragen hätte, wäre der neue Grad der persönlichen Überzeugung 94,1 Prozent. Und so weiter.

Die mathematischen Details sind in diesem Buch nicht relevant. Wir sollten uns nur zwei Dinge über die Bayessche Logik und unsere Neigung,

gegen sie zu verstoßen, merken. Erstens: Ungeachtet neuer Informationen über den konkreten Fall sind die Basisraten von Bedeutung. Dies ist oftmals nicht intuitiv einleuchtend. Zweitens: Intuitive Eindrücke über die Aussagekraft von Informationen sind oftmals übertrieben. Die Verknüpfung von WYSIATI und assoziativer Kohärenz lässt uns tendenziell die Geschichten glauben, die wir uns selbst ausdenken. Die wichtigsten Schlüssel zu einem disziplinierten bayesschen Denken lassen sich folgendermaßen zusammenfassen:

– Verankern Sie Ihr Urteil über die Wahrscheinlichkeit eines Ergebnisses in einer plausiblen Basisrate.
– Hinterfragen Sie die Aussagekraft Ihrer Informationen.

Beide Regeln sind einfach. Ich war betroffen, als ich erkannte, dass mir nie beigebracht worden war, wie man sie anwendet, und dass es mir selbst heute noch nicht in Fleisch und Blut übergegangen ist.

## Zum Thema »Repräsentativität«

»Der Rasen ist ordentlich geschnitten, die Rezeptionistin macht einen kompetenten Eindruck, und das Mobiliar ist geschmackvoll, aber dies bedeutet nicht, dass es ein gut gemanagtes Unternehmen ist. Ich hoffe, der Vorstand lässt sich nicht von Repräsentativität leiten.«

»Es sieht so aus, als könnte dieses neu gegründete Unternehmen gar nicht pleitegehen, aber die Basisraten des Erfolgs in dieser Branche sind extrem niedrig. Woher wissen wir, dass es in diesem Fall anders ist?«

»Sie machen immer wieder den gleichen Fehler: Auf einer dürftigen Datenbasis sagen sie seltene Ereignisse vorher. Wenn

die Datenbasis schwach ist, sollte man sich an die Basisraten halten.«

»Ich weiß, dass dieser Bericht absolut vernichtend ist, und er mag auf soliden Erkenntnissen basieren, aber wie sicher sind wir? Wir müssen diese Unsicherheit in unserem Denken mit einkalkulieren.«

# *15. Linda: Weniger ist mehr*

Das bekannteste und umstrittenste unserer Experimente drehte sich um eine fiktive Person namens Linda. Amos und ich dachten uns das Linda-Problem aus, um schlüssige Belege für die Rolle von Heuristiken in Urteilsprozessen und ihre Unvereinbarkeit mit den Regeln der Logik vorzulegen.[1] Wir beschrieben Linda folgendermaßen:

> Linda ist 31 Jahre alt, Single, freimütig und sehr intelligent. Sie hat Philosophie im Hauptfach studiert. Als Studentin interessierte sie sich sehr für Themen wie Diskriminierung und soziale Gerechtigkeit, und sie nahm auch an Anti-Atomkraft-Protesten teil.

Die Zuhörer, die diese Beschreibung in den 1980er-Jahren hörten, lachten immer, weil sie sofort wussten, dass Linda an der Universität von Kalifornien in Berkeley studiert hatte, die damals bekannt war für ihre radikalen, politisch engagierten Studenten. In einem unserer Experimente legten wir den Probanden eine Liste mit acht möglichen Szenarien für Linda vor. Wie beim Tom-W.-Problem stuften manche die Szenarien nach ihrer Repräsentativität ein, während andere sie nach ihrer Wahrscheinlichkeit ordneten. Das Linda-Problem ist ähnlich, allerdings mit einer besonderen Wendung.

> Linda ist Grundschullehrerin.
> Linda arbeitet in einer Buchhandlung und nimmt Yogakurse.
> Linda ist in der feministischen Bewegung aktiv.
> Linda betreut als Sozialarbeiterin psychisch Kranke.
> Linda ist Mitglied der League of Women Voters.
> Linda ist Bankkassiererin.

Linda ist Versicherungsvertreterin.
Linda ist Bankkassiererin und in der feministischen Bewegung aktiv.

Das Problem zeigt in mehrfacher Hinsicht einen zeitlichen Bezug. Die League of Women Voters spielt heute keine so bedeutende Rolle mehr, und die Idee einer »feministischen Bewegung« hört sich seltsam an – ein Beleg für den Wandel des Status der Frau in den letzten dreißig Jahren. Selbst im Zeitalter von Facebook lässt sich der fast perfekte Konsens der Urteile leicht erraten: Die Beschreibung von Linda passt sehr gut zu einer aktiven Feministin, recht gut zu einer Person, die in einer Buchhandlung arbeitet und Yogakurse besucht – und sehr schlecht zu einer Bankkassiererin oder einer Versicherungsvertreterin.

Konzentrieren wir uns jetzt auf die entscheidenden Elemente in der Liste: Passt die Beschreibung Lindas eher zu einer Bankkassiererin oder zu einer Bankkassiererin, die in der feministischen Bewegung aktiv ist? Alle sind sich einig, dass Linda eher der Vorstellung einer »feministischen Bankkassiererin« als dem Stereotyp einer Bankkassiererin entspricht. Die stereotype Bankkassiererin ist keine feministische Aktivistin, und wenn man die Beschreibung um dieses Detail ergänzt, erhält man eine kohärente Geschichte.

Die Besonderheit liegt in den Wahrscheinlichkeitsurteilen, weil zwischen den beiden Szenarien ein logischer Zusammenhang besteht. Denken Sie in Venn-Diagrammen (Mengendiagrammen). Die Menge feministischer Bankkassiererinnen ist vollständig in der Menge der Bankkassiererinnen enthalten, da jede feministische Bankkassiererin eine Bankkassiererin ist. Daher *muss* die Wahrscheinlichkeit, dass Linda eine feministische Bankkassiererin ist, niedriger sein als die Wahrscheinlichkeit, dass sie eine Bankkassiererin ist. Wenn man ein potenzielles Ereignis detaillierter beschreibt, kann man seine Wahrscheinlichkeit nur vermindern. Das Problem löst daher einen Konflikt zwischen der Intuition der Repräsentativität und der Logik der Wahrscheinlichkeit aus.

Bei unserem ersten Experiment hielten wir uns an das sogenannte »Between-Subjects-Design«, das heißt, wir variierten die experimentellen Bedingungen zwischen zwei gesonderten Probandengruppen. Alle Probanden sahen eine Reihe von sieben Ergebnissen, die jeweils nur eines

der kritischen Elemente (»Bankkassiererin« oder »feministische Bankkassiererin«) beinhalteten. Einige stuften die Ergebnisse nach Ähnlichkeit ein, andere nach Wahrscheinlichkeit. Wie im Fall von Tom W. waren die durchschnittlichen Einstufungen nach Ähnlichkeit und nach Wahrscheinlichkeit identisch; »feministische Bankkassiererinnen« wurden in beiden höher eingestuft als »Bankkassiererinnen«.

Dann verfeinerten wir das Experiment mit einem »Within-Subject-Design«, das heißt, wir variierten die experimentellen Bedingungen innerhalb einer Probandengruppe. Wir konzipierten den Fragebogen so, wie Sie ihn gesehen haben, mit »Bankkassiererin« an der sechsten Stelle der Liste und »feministische Bankkassiererin« als letztem Element. Wir waren überzeugt davon, dass die Probanden die Beziehung zwischen den beiden Ergebnissen bemerken würden und dass ihre Einstufungen konsistent mit der statistischen Logik wären. Tatsächlich waren wir uns dessen so sicher, dass wir es nicht für notwendig erachteten, ein gesondertes Experiment durchzuführen. Meine Assistentin führte im Labor ein anderes Experiment durch, und sie bat die Versuchspersonen, den neuen Linda-Fragebogen auszufüllen, während sie sich abmeldeten, unmittelbar bevor sie ihr Honorar erhielten.

Etwa zehn Fragebogen hatten sich in einem Ablagekorb auf dem Schreibtisch meiner Assistentin angesammelt, als ich beiläufig einen Blick darauf warf und bemerkte, dass alle Probanden »feministische Bankkassiererin« für wahrscheinlicher hielten als »Bankkassiererin«. Ich war so überrascht, dass ich noch immer eine »Blitzlichterinnerung« an die graue Farbe des Metallschreibtisches und an die Standorte aller Anwesenden habe, als ich diese Entdeckung machte. Ich rief Amos sofort total begeistert an, um ihm mitzuteilen, was ich gefunden hatte: Wir hatten die Logik gegen die Repräsentativität ausgespielt, und die Repräsentativität hatte gewonnen!

Wir hatten, um in der Begrifflichkeit dieses Buches zu bleiben, ein Versagen von System 2 beobachtet: Unsere Probanden hatten eine ordentliche Chance, die Relevanz der logischen Regel zu erkennen, da beide Ergebnisse im selben Ranking enthalten waren. Aber sie haben diese Chance nicht genutzt. Als wir das Experiment erweiterten, stellten wir fest, dass 89 Prozent der Studenten in unserer Stichprobe gegen die Logik der Wahrscheinlichkeit verstießen. Da wir überzeugt davon waren, dass Probanden mit Statistikkenntnissen besser abschneiden würden, teilten wir

denselben Fragebogen an Doktoranden des Studiengangs Entscheidungswissenschaft der Stanford Graduate School of Business aus, die alle mehrere vertiefende Lehrveranstaltungen in Wahrscheinlichkeits- und Entscheidungstheorie sowie Statistik besucht hatten. Wir waren wieder überrascht: 85 Prozent dieser Probanden hielten »feministische Bankkassiererin« für wahrscheinlicher als »Bankkassiererin«.

In »zunehmend verzweifelten« Bemühungen, den Fehler zu beseitigen, wie wir später schrieben, machten wir große Probandengruppen mit Linda bekannt und stellten ihnen diese einfache Frage:

> Welche Alternative ist wahrscheinlicher?
> Linda ist eine Bankkassiererin?
> Linda ist eine Bankkassiererin und in der feministischen Bewegung aktiv?

Diese starke Version des Problems machte Linda in manchen Kreisen berühmt, und sie trug uns eine jahrelang Kontroverse ein. Etwa 85 bis 90 Prozent der Studenten an mehreren großen amerikanischen Universitäten entschieden sich – entgegen der Logik der Wahrscheinlichkeit – für die zweite Option. Bemerkenswerterweise schienen die Sünder keine Scham zu empfinden. Als ich die Teilnehmer eines großen Seminars fragte: »Ist Ihnen bewusst, dass Sie gegen eine elementare logische Regel verstoßen haben?«, rief jemand in der letzten Reihe: »Na und?« Und eine Studentin, die denselben Fehler gemacht hatte, rechtfertigte sich mit der Aussage: »Ich dachte, Sie hätten mich nur nach meiner Meinung gefragt.«

Das Wort »Fehlschluss« wird im Allgemeinen dann benutzt, wenn Menschen eine logische Regel, die offenkundig relevant ist, nicht anwenden. Amos und ich führten das Konzept des »Konjunktionsfehlschlusses« ein, den Menschen begehen, wenn sie eine Verknüpfung zweier Ereignisse (hier Bankkassiererin und Feministin) im direkten Vergleich als wahrscheinlicher beurteilen als eines der Ereignisse (Bankkassiererin).

Wie bei der Müller-Lyer-Illusion bleibt der Fehlschluss auch dann verlockend, wenn man ihn als solchen erkennt. Der Evolutionsforscher Stephen Jay Gould beschrieb sein eigenes Ringen mit dem Linda-Problem.

Er kannte natürlich die richtige Antwort, und dennoch schrieb er: »Ein kleiner Homunkulus in meinem Kopf springt weiterhin auf und ab und schreit mich an: ›Aber sie kann nicht nur eine Bankkassiererin sein; lies die Beschreibung.‹«[2] Der kleine Homunkulus ist natürlich Goulds System 1, das eindringlich auf ihn einredet. (Die Zwei-System-Terminologie war zu dem Zeitpunkt, als er schrieb, noch nicht eingeführt.)

Die richtige Antwort auf diese Kurzfassung des Linda-Problems war nur in einer unserer Studien die Mehrheitsantwort: 64 Prozent einer Gruppe von Graduierten der Sozialwissenschaften an den Universitäten Stanford und Berkeley urteilten zutreffend, dass »feministische Bankkassiererin« weniger wahrscheinlich als »Bankkassiererin« ist. In der ursprünglichen Version mit acht Ergebnissen (wie weiter oben gezeigt) hatten nur 15 Prozent einer ähnlichen Gruppe von vorgerückten Studenten diese Wahl getroffen. Der Unterschied ist lehrreich. Die längere Version trennte die beiden entscheidenden Ergebnisse durch ein eingeschobenes Element (»Versicherungsvertreterin«), und die Leser beurteilten jedes Ergebnis unabhängig von den anderen, ohne sie miteinander zu vergleichen. Die kürzere Version dagegen erforderte einen expliziten Vergleich, der System 2 mobilisierte und den meisten Studenten mit Statistikkenntnissen erlaubte, den Fehlschluss zu vermeiden. Leider haben wir die Gedankengänge der erheblichen Minderheit (36 Prozent) in dieser sachkundigen Probandengruppe, die falsch entschieden hat, nicht weiter untersucht.

Die Wahrscheinlichkeitsurteile unserer Probanden beim Tom-W.- und beim Linda-Problem entsprachen genau den Repräsentativitätsurteilen (Ähnlichkeit mit Stereotypen). Repräsentativität gehört zu einem Cluster eng miteinander verbundener grundlegender Bewertungen, die mit hoher Wahrscheinlichkeit gemeinsam generiert werden. Die repräsentativsten Ergebnisse führen in Verbindung mit der Persönlichkeitsbeschreibung zu den kohärentesten Geschichten. Die kohärentesten Geschichten sind nicht unbedingt die wahrscheinlichsten, aber sie sind *plausibel*, und die Konzepte der Kohärenz, Plausibilität und Wahrscheinlichkeit werden von den Unbesonnenen leicht durcheinandergebracht.

Die unkritische Ersetzung der Wahrscheinlichkeit durch Plausibilität wirkt sich schädlich auf Urteile aus, wenn Szenarien als Instrumente der Prognose benutzt werden. Betrachten Sie die folgenden beiden Szenarien, die verschiedenen Gruppen mit der Forderung vorgelegt wurden, ihre Wahrscheinlichkeit einzuschätzen:

Eine gewaltige Flutkatastrophe irgendwo in Nordamerika im nächsten Jahr, bei der mehr als tausend Menschen ertrinken.

Ein Erdbeben in Kalifornien irgendwann im nächsten Jahr, das eine Flutkatastrophe verursacht, bei der mehr als tausend Menschen ertrinken.

Das Szenario eines Erdbebens in Kalifornien ist plausibler als das nordamerikanische Szenario, auch wenn seine Wahrscheinlichkeit mit Sicherheit geringer ist. Erwartungsgemäß waren die Wahrscheinlichkeitseinschätzungen für das informativere und detailliertere Szenario höher, obwohl dies der Logik widersprach. Das ist die Falle für Prognostiker und ihre Kunden: Je detailreicher Szenarien sind, umso überzeugender sind sie, aber desto geringer ist ihre Wahrscheinlichkeit. Um zu verstehen, welche Rolle die Plausibilität spielt, betrachten Sie die folgenden Fragen:

Welche Alternative ist wahrscheinlicher?

Mark hat Haare.

Mark hat blondes Haar.

Und:

Welche Alternative ist wahrscheinlicher?

Jane ist Lehrerin.

Jane ist Lehrerin und geht zu Fuß zur Arbeit.

Die beiden Fragen haben die gleiche logische Struktur wie das Linda-Problem, aber sie erzeugen keinen Fehlschluss, weil die detailliertere Aussage nur detaillierter ist – nicht plausibler oder kohärenter oder eine bessere Geschichte. Die Beurteilung von Plausibilität und Kohärenz beantwortet die Wahrscheinlichkeitsfrage nicht. Die Logik setzt sich durch, wenn keine konkurrierenden intuitiven Einschätzungen vorhanden sind.

## Weniger ist mehr, manchmal sogar bei gemeinsamer Bewertung

Christopher Hsee von der Universität Chicago bat Personen, Preise für Geschirrservice festzusetzen, die im Rahmen eines Ausverkaufs in einem örtlichen Geschäft angeboten wurden, wo ein Satz Tafelgeschirr regulär zwischen 30 und 60 Dollar kostet. In diesem Experiment gab es drei Gruppen. Die unten stehende Übersicht wurde einer Gruppe gezeigt; Hsees bezeichnet dies als »gemeinsame Bewertung«, weil sie einen Vergleich der beiden Service erlaubt. Den anderen beiden Gruppen wurde eines der beiden Service gezeigt; dies ist eine »einfache Bewertung«. Die gemeinsame Bewertung ist ein Experiment mit Within-Subject-Design, während die einfache Bewertung ein Between-Subjects-Design hat.

| | Service A: 40 Einzelteile | Service B: 24 Einzelteile |
|---|---|---|
| Große Teller | 8, alle in gutem Zustand | 8, alle in gutem Zustand |
| Suppen-/Salatteller | 8, alle in gutem Zustand | 8, alle in gutem Zustand |
| Dessertteller | 8, alle in gutem Zustand | 8, alle in gutem Zustand |
| Tassen | 8, 2 zerbrochen | |
| Untertassen | 8, 7 zerbrochen | |

Angenommen, beide Service sind von gleicher Qualität, welches ist dann mehr wert? Die Frage ist leicht. Sie können sehen, dass Service A sämtliche

Einzelteile von Service B – und sieben weitere unbeschädigte Teile – enthält, daher *muss* es mehr wert sein. Tatsächlich waren die Teilnehmer an Hsees Experiment mit gemeinsamer Bewertung bereit, ein wenig mehr für Service A als für Service B zu bezahlen: 32 gegenüber 30 Dollar.

Bei der Einzelbewertung war es genau umgekehrt: Für Service B wurde ein viel höherer Preis festgesetzt als für Service A: 33 gegenüber 23 Dollar. Wir wissen, wieso dies geschah. Mengen (und dazu gehören auch Geschirrservice) werden durch Normen und Prototypen repräsentiert. Sie spüren sofort, dass der durchschnittliche Wert der Einzelteile in Service A viel niedriger ist als in Service B, weil niemand für zerbrochenes Geschirr zahlen will. Wenn der Durchschnitt maßgeblich für die Bewertung ist, ist es nicht überraschend, dass der Wert von Service B höher veranschlagt wird. Hsee nannte das so entstandene Muster »weniger ist mehr«. Die Entfernung von 16 Einzelteilen aus Service A (darunter sieben unbeschädigten) steigerte dessen Wert.

Hsees Ergebnisse wurden von dem Experimentalökonomen John List in einem realen Markt für Baseball-Eintrittskarten bestätigt. Er versteigerte Sätze von jeweils zehn hochwertigen Karten und identische Sätze, zu denen drei geringwertige Karten dazukamen. Wie in dem Service-Experiment wurden die größeren Kartensätze bei der gemeinsamen Bewertung höher bewertet als die kleineren Sätze. Bei der Einzelbewertung dagegen wurden sie niedriger bewertet. Aus wirtschaftstheoretischer Perspektive ist dies ein verstörendes Ergebnis: Der ökonomische Wert eines Geschirrservice oder einer Sammlung Baseball-Eintrittskarten ist eine summenähnliche Variable. Wenn man ein positiv bewertetes Element zu der Summe addiert, nimmt deren Wert zwangsläufig zu.

Das Linda-Problem und das Service-Problem haben genau die gleiche Struktur. Die Wahrscheinlichkeit ist, wie der ökonomische Wert, eine summenähnliche Variable, wie dieses Beispiel veranschaulicht:

Wahrscheinlichkeit (Linda ist eine Bankkassiererin)
= Wahrscheinlichkeit (Linda ist eine feministische Bankkassiererin)

+ Wahrscheinlichkeit (Linda ist eine nicht feministische Bankkassiererin)

Dies ist auch der Grund dafür, dass Einzelbeurteilungen des Linda-Problems ein »Weniger-ist-mehr-Muster« erzeugen, wie es in Hsees Geschirrservice-Studie der Fall war. System 1 mittelt, statt zu addieren, sodass die subjektive Wahrscheinlichkeit zunimmt, wenn die nicht feministischen Bankkassiererinnen aus der Menge entfernt werden. Doch die summenähnliche Natur der Variablen ist für Wahrscheinlichkeiten weniger offensichtlich als für Geld. Infolgedessen beseitigt die gemeinsame Bewertung den Fehler nur in Hsees Experiment, nicht im Linda-Experiment.

Linda war nicht der einzige Verknüpfungsfehler, der die gemeinsame Bewertung überlebte. Bei vielen anderen Urteilen fanden wir ähnliche Verstöße gegen die Logik. Die Probanden in einer dieser Studien sollten vier mögliche Ergebnisse des nächsten Wimbledon-Turniers nach ihrer Wahrscheinlichkeit einstufen. Zur damaligen Zeit war Björn Borg der erfolgreichste Tennisspieler. Dies waren die Ergebnisse:

a. Borg gewinnt das Spiel.
b. Borg verliert den ersten Satz.
c. Borg verliert den ersten Satz, aber gewinnt das Spiel.
d. Borg gewinnt den ersten Satz, aber verliert das Spiel.

Die kritischen Aussagen sind B und C. B ist das umfassendere Ereignis, und seine Wahrscheinlichkeit *muss* höher sein als die eines Ereignisses, das in ihm enthalten ist. Im Widerspruch zur Logik, aber nicht zur Repräsentativität oder Plausibilität schrieben 72 Prozent B eine niedrigere Wahrscheinlichkeit zu als C – ein weiterer Fall, wo bei einem direkten Vergleich weniger mehr ist. Auch hier war das Szenario, das als wahrscheinlicher beurteilt wurde, fraglos plausibler, und es passte besser zu all dem, was man über den besten Tennisspieler der Welt wusste.

Um dem möglichen Einwand zu begegnen, der Konjunktionsfehlschluss sei auf eine Fehlinterpretation des Wortes »Wahrscheinlichkeit« zurückzuführen, konstruierten wir ein Problem, das Wahrscheinlichkeitsurteile erforderte, bei denen die Ereignisse jedoch nicht mit Wörtern beschrieben wurden, und in dem der Ausdruck

»Wahrscheinlichkeit« nicht auftauchte. Wir erzählten den Probanden von einem sechsseitigen Standardwürfel mit vier grünen und zwei roten Seiten, der zwanzigmal geworfen werden sollte. Wir zeigten ihnen drei Folgen von Grün (G) und Rot (R) und forderten sie auf, eine davon auszuwählen. Sie würden (hypothetisch) 25 Dollar gewinnen, wenn die von ihnen gewählte Folge gewürfelt würde. Die Folgen waren:

1. RGRRR
2. GRGRRR
3. GRRRRR

Da der Würfel doppelt so viele grüne wie rote Seiten hat, ist die erste Folge eher nicht repräsentativ – wie die Aussage, Linda sei eine Bankkassiererin. Die zweite Folge mit sechs Würfen passt besser zu dem, was wir von diesem Würfel erwarten würden, weil sie zweimal G enthält. Allerdings wurde diese Folge durch Anfügung eines G an den Anfang der ersten Folge generiert, sodass sie nur unwahrscheinlicher sein kann als die erste. Dies ist die nonverbale Entsprechung zu der Aussage, dass Linda eine feministische Bankkassiererin ist. Wie in der Linda-Studie war die Repräsentativität maßgeblich. Fast zwei Drittel der Probanden wetteten auf Folge 2 statt auf Folge 1. Doch als man ihnen Argumente für die beiden Optionen vorlegte, fand eine große Mehrheit das richtige Argument (das Folge 1 begünstigt) überzeugender.

Das nächste Problem war ein Durchbruch, weil wir endlich eine Bedingung fanden, in der die Häufigkeit des Konjunktionsfehlschlusses stark reduziert war. Zwei Probandengruppen sahen geringfügig voneinander abweichende Varianten desselben Problems:

| | |
|---|---|
| Eine Stichprobe erwachsener Männer aller Altersgruppen und Berufe in British Columbia wurde auf ihre gesundheitliche Verfassung untersucht. Bitte geben Sie Ihre beste Einschätzung für folgende Werte an: | Eine Stichprobe von hundert erwachsenen Männern aller Altersgruppen und Berufe in British Columbia wurde auf ihre gesundheitliche Verfassung untersucht. Bitte geben Sie Ihre beste Einschätzung für folgende Werte an: |
| Wie viel Prozent der untersuchten Männer hatten bereits einen oder mehrere Herzinfarkte? | Wie viele der hundert Teilnehmer hatten einen oder mehrere Herzinfarkte? |
| Wie viel Prozent der untersuchten Männer sind über 55 Jahre alt und hatten bereits einen oder mehrere Herzinfarkte? | Wie viele der hundert Teilnehmer sind über 55 Jahre alt und hatten einen oder mehrere Herzinfarkte? |

Die Fehlerhäufigkeit war 65 Prozent in der Gruppe, die das Problem in der linken Version sah, und 25 Prozent in der Gruppe, die das Problem in der rechten Version sah.

Wieso ist die Frage »Wie viele der hundert Teilnehmer …« so viel leichter als »Wie viel Prozent …«? Eine wahrscheinliche Erklärung besteht darin, dass die Bezugnahme auf hundert Personen eine räumliche Repräsentation im Gehirn aktiviert. Stellen Sie sich vor, eine große Zahl von Menschen wird aufgefordert, sich in einem Raum zu Gruppen zu ordnen: »Diejenigen, deren Namen mit den Buchstaben A bis L anfangen, werden aufgefordert, sich in der vorderen linken Ecke zu versammeln.« Anschließend sollen sie sich selbst weitersortieren. Die Beziehung der Inklusion ist jetzt offensichtlich, und man sieht, dass Personen, deren Namen mit C beginnen, eine Teilmenge der Menge in der vorderen linken Ecke bilden. In der Frage nach dem Gesundheitszustand landen diejenigen, die schon einmal einen Herzinfarkt erlitten, in einer Ecke des Raumes, und einige von ihnen sind jünger als 55 Jahre. Nicht jeder hat diese besonders anschaulichen Vorstellungsbilder, aber viele spätere Experimente haben gezeigt, dass sich mithilfe der sogenannten Häufigkeitsrepräsentation leicht erkennen lässt, dass eine Gruppe vollständig in der anderen enthalten ist.

Die Lösung des Rätsels scheint darin zu liegen, dass eine Frage, die als »Wie viele?« formuliert wird, uns an Einzelpersonen denken lässt, während die gleiche Frage mit der Formulierung »Wie viel Prozent?« dies nicht tut.

Was haben wir aus diesen Studien über die Funktionsmechanismen von System 2 gelernt? Eine Schlussfolgerung, die nicht neu ist, lautet, dass System 2 nicht besonders wach ist. Die Studenten, die an unseren Studien über den Konjunktionsfehlschluss teilnahmen, »kannten« zweifellos die Logik von Venn-Diagrammen, aber sie haben sie nicht in zuverlässiger Weise angewandt, auch dann nicht, wenn alle relevanten Informationen vor ihnen ausgebreitet wurden. Die Absurdität des Weniger-ist-mehr-Musters zeigte sich deutlich in Hsees Geschirrservice-Studie, und sie ließ sich leicht in der »Wie-viele-Repräsentation« erkennen. Aber für die Tausende von Probanden, die bei dem ursprünglichen Linda-Problem und bei anderen, ähnlichen Problemen den Konjunktionsfehlschluss begingen, war diese Absurdität nicht offensichtlich. In all diesen Fällen schien die Konjunktion plausibel zu sein, und dies genügte für eine Unterstützung von System 2.

Die Faulheit von System 2 ist Teil der Geschichte. Wenn der nächste Urlaub unserer Probanden davon abhängig gewesen wäre und wenn man ihnen unbegrenzt Zeit gegeben und sie aufgefordert hätte, sich an die Logik zu halten und erst dann zu antworten, wenn sie ihrer Antwort sicher seien, dann, so glaube ich, hätten die meisten unserer Probanden den Konjunktionsfehlschluss vermieden. Aber ihr Urlaub hing eben nicht von der richtigen Antwort ab; sie wandten sehr wenig Zeit für die Aufgabe auf und begnügten sich damit, so zu antworten, als wären sie nur »nach ihrer Meinung gefragt worden«. Die Faulheit von System 2 ist eine wichtige Tatsache unseres geistigen Lebens, und die Beobachtung, dass Repräsentativität die Anwendung einer naheliegenden logischen Regel blockieren kann, ist ebenfalls nicht ganz belanglos.

Das Bemerkenswerte an der Linda-Geschichte ist der Gegensatz zur Studie mit dem zerbrochenen Geschirr. Die beiden Probleme haben die gleiche Struktur, führen aber zu unterschiedlichen Ergebnissen. Probanden, die die Information erhalten, dass das Service zerbrochene Teile enthält, setzen einen sehr niedrigen Preis dafür fest; in ihrem Verhalten spiegelt sich eine intuitive Regel wider. Andere, die Informationen über beide Service gleichzeitig erhalten haben, wenden die logische Regel an, dass mehr Geschirrteile den Wert nur steigern können. Die Intuition prägt die Urteile in der experimentellen Bedingung des Between-Subjects-Designs (mit

getrennten Probandengruppen); die Logik gibt den Ausschlag bei gemeinsamer Beurteilung. Beim Linda-Problem dagegen hat die Intuition selbst bei gemeinsamer Bewertung oftmals die Logik überwunden, auch wenn wir einige Bedingungen identifizierten, in denen sich die Logik durchsetzte.

Amos und ich waren der Ansicht, dass die eklatanten Verstöße gegen die Logik der Wahrscheinlichkeit, die wir bei transparenten Problemen beobachtet hatten, interessant waren und es verdienten, unseren Kollegen mitgeteilt zu werden. Wir glaubten auch, dass die Ergebnisse unser Argument über die Wirkmächtigkeit von Urteilsheuristiken untermauerten und dass sie auch Zweifler überzeugen würden. Aber darin sollten wir uns gründlich irren. Stattdessen wurde das Linda-Problem zu einer Fallstudie über die Normen wissenschaftlicher Kontroversen.

Das Linda-Problem erregte viel Aufmerksamkeit, aber es wurde auch zu einem Magneten für Kritiker unserer Herangehensweise an die Psychologie von Urteilsprozessen. Forscher fanden Kombinationen von Instruktionen und Hinweisen, die die Häufigkeit des Fehlschlusses verringerten – was wir selbst bereits getan hatten; einige behaupteten, im Kontext des Linda-Problems sei es aus Sicht der Probanden nur vernünftig, das Wort »Wahrscheinlichkeit« im Sinne von »Plausibilität« zu interpretieren. Diese Argumente wurden manchmal erweitert zu der Behauptung, unser ganzes Unterfangen sei verfehlt: Wenn eine so markante (saliente) kognitive Illusion abgeschwächt oder wegerklärt werden könne, dann gelte dies erst recht für andere.[3] Diese Argumentation vernachlässigt das einzigartige Kennzeichen des Konjunktionsfehlschlusses als Konflikt zwischen Intuition und Logik. Die empirischen Belege, die wir bei Between-Subjects-Experimenten (einschließlich Studien über Linda) für Heuristiken zusammengetragen hatten, wurden nicht infrage gestellt – sie wurden gar nicht angesprochen, und ihre Relevanz wurde durch die ausschließliche Konzentration auf den Konjunktionsfehler verringert. Das Linda-Problem führte letztlich dazu, dass unsere Arbeit in der Öffentlichkeit verstärkt wahrgenommen wurde, während die Glaubwürdigkeit unseres Ansatzes bei unseren Fachkollegen ein wenig litt. Damit hatten wir überhaupt nicht gerechnet.

Wenn Sie eine Gerichtsverhandlung besuchen, wird Ihnen auffallen, dass Rechtsanwälte zwei Arten von Kritik vorbringen: Um eine Klage zum Scheitern zu bringen, ziehen sie die stärksten Argumente, auf denen diese

beruht, in Zweifel; um einen Zeugen zu diskreditieren, konzentrieren sie sich auf den schwächsten Teil seiner Aussage. Auch in politischen Debatten konzentrieren sich die Kontrahenten normalerweise auf die Schwächen. In wissenschaftlichen Kontroversen halte ich das nicht für angemessen, aber ich habe mich mit der Tatsache abgefunden, dass die Normen der Debatte in den Sozialwissenschaften eine »politische« Streitführung nicht verbieten, insbesondere bei weitreichenden Fragen – und die Häufigkeit von Urteilsfehlern ist eine weitreichende Frage.

Vor einigen Jahren hatte ich ein freundschaftliches Gespräch mit Ralph Hertwig, einem beharrlichen Kritiker des Linda-Problems, mit dem ich in dem vergeblichen Bemühen, unsere Differenzen beizulegen, zusammengearbeitet hatte.[4] Ich fragte ihn, wieso er und andere sich ausschließlich auf den Konjunktionsfehlschluss konzentrierten statt auf andere Befunde, die unsere Position stärker unterstützten. Er antwortete lächelnd: »Es war interessanter.« Und er fügte hinzu, das Linda-Problem habe so viel Beachtung gefunden, dass wir keinen Grund zur Klage hätten.

## Zum Thema »Weniger ist mehr«

»Sie dachten sich ein sehr kompliziertes Szenario aus und bestanden darauf, es als hochwahrscheinlich zu beurteilen. Das ist es nicht – es ist nur eine plausible Geschichte.«

»Sie fügten dem teuren Produkt ein billiges Geschenk bei und machten das ganze Geschäft dadurch weniger attraktiv. In diesem Fall ist weniger mehr.«

»In den meisten Situationen macht ein direkter Vergleich Menschen bedächtiger und logischer. Aber nicht immer.

Manchmal sticht die Intuition die Logik auch dann aus, wenn einem die richtige Antwort in die Augen springt.«

# *16. Ursachen vs. Statistik*

Betrachten Sie das folgende Szenario und achten Sie auf Ihre intuitive Antwort auf die Frage.

> Ein Taxi war an einem nächtlichen Unfall mit Fahrerflucht beteiligt.
> Zwei Taxi-Unternehmen, Green und Blue, arbeiten in der Stadt.
> Sie erhalten die folgenden Daten:
>
> – 85 Prozent der Taxis in der Stadt gehören der Firma Green, 15 Prozent der Firma Blue.
> – Ein Zeuge hat das Taxi als eines der Firma Blue identifiziert. Das Gericht überprüfte die Zuverlässigkeit des Zeugen unter den Umständen, die in der Nacht des Unfalls herrschten, und gelangte zu dem Schluss, dass der Zeuge jede der beiden Farben in 80 Prozent der Fälle richtig und in 20 Prozent der Fälle nicht richtig erkannte.
>
> Wie hoch ist die Wahrscheinlichkeit, dass das an dem Unfall beteiligte Taxi der Firma Blue und nicht der Firma Green gehörte?

Dies ist ein Standardproblem der Bayesschen Inferenz. Es gibt zwei Informationselemente: eine Basisrate und die nicht gänzlich zuverlässige Aussage eines Zeugen. Ohne einen Zeugen liegt die Wahrscheinlichkeit dafür, dass das am Unfall beteiligte Taxi der Firma Blue gehört, bei 15 Prozent – was die Basisrate für dieses Ergebnis ist. Wenn die beiden Taxi-Unternehmen gleich groß wären, hätte die Basisrate keinen Informationsgehalt, und Sie würden lediglich die Zuverlässigkeit des Zeugen betrachten, sodass Sie zu dem Schluss kämen, dass die Wahrscheinlichkeit 80 Prozent beträgt. Die beiden Informationsquellen

können gemäß der Bayesschen Regel miteinander kombiniert werden. Die richtige Antwort lautet 41 Prozent.[1] Doch vermutlich ahnen Sie, was die meisten Menschen tun, wenn sie mit diesem Problem konfrontiert sind: Sie ignorieren die Basisrate und halten sich an den Zeugen. Die häufigste Antwort lautet 80 Prozent.

## Kausale Stereotype

Betrachten wir nun eine Variante der gleichen Geschichte, in der nur die Präsentation der Basisrate verändert wurde.

Sie erhalten die folgenden Daten:

– Die beiden Firmen betreiben die gleiche Anzahl von Taxis, aber Green-Taxis sind an 85 Prozent der Unfälle beteiligt.
– Die Information über den Zeugen ist die gleiche wie in der vorangehenden Version.

Die beiden Versionen des Problems sind mathematisch gesehen identisch, aber psychologisch betrachtet recht unterschiedlich. Personen, die die erste Version lesen, wissen nichts mit der Basisrate anzufangen und ignorieren sie oftmals. Dagegen geben die Probanden, denen die zweite Version dargeboten wird, der Basisrate ein erhebliches Gewicht, und ihr durchschnittliches Urteil ist nicht allzu weit von der bayesschen Lösung entfernt.[2] Wieso?

In der ersten Version ist die Basisrate von Blue-Taxis eine statistische Tatsache über die Taxis in der Stadt. Ein auf kausale Zusammenhänge erpichter Intellekt findet hier nichts zum Beißen: Wie kann die Anzahl der Green- und Blue-Taxis in der Stadt die Ursache dafür sein, dass dieser Taxifahrer Unfallflucht begeht?

In der zweiten Version dagegen verursachen die Fahrer der Green-Taxis mehr als fünfmal so viele Unfälle wie die Blue-Taxis. Daher die prompte Schlussfolgerung: Die Green-Fahrer müssen ein Verein von rücksichtslosen Irren sein! Sie haben jetzt ein Stereotyp über die Rücksichtslosigkeit von Green gebildet, das Sie auf unbekannte einzelne Fahrer im Unternehmen anwenden. Das Stereotyp lässt sich mühelos in eine kausale Geschichte einfügen, weil Rücksichtslosigkeit eine kausal relevante Information über einzelne Taxifahrer ist. Diese Version enthält zwei kausale Geschichten, die miteinander kombiniert beziehungsweise in Einklang gebracht werden müssen. Die erste ist die Fahrerflucht, die natürlich die Vorstellung wachruft, dass ein rücksichtsloser Green-Fahrer verantwortlich war. Die zweite ist die Zeugenaussage, die entschieden darauf hindeutet, dass es ein Blue-Taxi war. Die Schlüsse aus den beiden Geschichten über die Farbe des

Taxis sind widersprüchlich und heben sich annähernd gegenseitig auf. Die Wahrscheinlichkeit der beiden Farben ist etwa gleich groß (der Bayessche Schätzwert beträgt 41 Prozent, worin sich die Tatsache widerspiegelt, dass die Basisrate von Green-Taxis etwas größer ist als die Zuverlässigkeit des Zeugen, der ein blaues Taxi gesehen zu haben glaubt).

Das Taxi-Beispiel veranschaulicht zwei Typen von Basisraten. »Statistische Basisraten« sind Tatsachen über eine Population, zu der ein Einzelfall gehört, aber sie sind nicht relevant für den Einzelfall. »Kausale Basisraten« verändern Ihre Sichtweise darüber, was einen Einzelfall hervorgebracht hat. Die beiden Typen von Informationen über Basisraten werden unterschiedlich behandelt:

- Statistische Basisraten werden im Allgemeinen untergewichtet und manchmal völlig außer Acht gelassen, wenn spezifische Informationen über den vorliegenden Fall verfügbar sind.
- Kausale Basisraten werden als Informationen über den Einzelfall behandelt und lassen sich leicht mit anderen fallspezifischen Informationen verknüpfen.

Die kausale Version des Taxi-Problems hatte die Form eines Stereotyps: Green-Fahrer sind gefährlich. Stereotype sind Aussagen über die Gruppe, die (zumindest vorläufig) als Tatsachen über jedes Mitglied dieser Gruppe akzeptiert werden. Hier sind zwei Beispiele:

> Die meisten Absolventen dieser innerstädtischen Schule besuchen das College.
> Das Interesse an Radsport ist in Frankreich weit verbreitet.

Diese Aussagen werden bereitwillig so interpretiert, dass sie eine Verhaltenstendenz bei einzelnen Mitgliedern der Gruppe auslösen, und sie passen in eine kausale Geschichte. Viele Absolventen dieser konkreten innerstädtischen Schule sind bestrebt und befähigt, aufs College zu gehen, vermutlich aufgrund eines förderlichen Lernumfeldes in dieser Schule. Es gibt Kräfte in der französischen Kultur und Gesellschaft, die viele Franzosen dazu veranlassen, sich für Radsport zu interessieren. Diese Tatsachen werden Ihnen in Erinnerung gerufen, wenn Sie über die

Wahrscheinlichkeit nachdenken, mit der ein bestimmter Schulabsolvent aufs College gehen wird, oder wenn Sie sich fragen, ob Sie die Tour de France im Gespräch mit einem Franzosen, den Sie gerade kennengelernt haben, ansprechen sollen.

»Stereotypisierung« ist in unserer Kultur ein negativ besetztes Wort, aber ich verwende es in einem neutralen Sinn. Eines der grundlegenden Merkmale von System 1 besteht darin, dass es Kategorien als Normen und mustergültige Beispiele repräsentiert. So stellen wir uns Pferde, Kühlschränke und New Yorker Polizisten vor; wir halten eine Repräsentation eines oder mehrerer »normaler« Mitglieder von jeder dieser Kategorien im Gedächtnis. Wenn es sich um soziale Kategorien handelt, werden die Repräsentationen Stereotype genannt. Einige Stereotype sind in schädlicher Weise falsch, und feindseliges Stereotypisieren kann schreckliche Folgen haben, aber die psychologischen Tatsachen lassen sich nicht ignorieren: Wir stellen uns Kategorien als Stereotype vor, die sowohl richtig als auch falsch sein können.

Vielleicht bemerken Sie die Ironie. Im Zusammenhang mit dem Taxi-Problem ist die Vernachlässigung der Basisraten-Information eine kognitive Fehlleistung, ein Verstoß gegen die Bayessche Logik, und es ist wünschenswert, hier kausale Basisraten heranzuziehen. Die Stereotypisierung der Green-Fahrer verbessert die Urteilsgenauigkeit. Doch in anderen Zusammenhängen, wie etwa bei der Personaleinstellung oder beim Erstellen von Persönlichkeitsprofilen, widerstreben starke soziale Normen einer Stereotypisierung, und diese sind auch gesetzlich verankert. Und so sollte es auch sein. In sensiblen sozialen Kontexten wollen wir auf der Basis statistischer Gruppendaten keine potenziell fehlerhaften Schlussfolgerungen über eine Einzelperson ziehen. Wir erachten es als moralisch wünschenswert, dass Basisraten als statistische Tatsachen über die Gruppe behandelt werden statt als mutmaßliche Fakten über Einzelpersonen. Anders gesagt, wir lehnen kausale Basisraten ab.

Die soziale Norm gegen Stereotypisierung, wozu auch die Ablehnung von Persönlichkeitsprofilen zählt, war beim Aufbau einer zivilisierteren Gesellschaft mit größerer sozialer Gleichheit überaus nützlich. Man sollte sich trotzdem daran erinnern, dass die Vernachlässigung gültiger Stereotype zwangsläufig zu suboptimalen Urteilen führt. Widerstand gegen Stereotypenbildung ist ein löblicher moralischer Standpunkt, aber die

Vorstellung, der Widerstand wäre kostenlos, ist falsch. Die Kosten lohnen sich, um eine bessere Gesellschaft zu verwirklichen, aber zu bestreiten, dass die Kosten existieren, mag unserem Gewissen schmeicheln und politisch korrekt sein, ist jedoch wissenschaftlich nicht haltbar. In emotional aufgeladenen politischen Kontroversen ist es üblich, sich auf Affektheuristiken zu verlassen. Die Standpunkte, die wir vertreten, verursachen keine Kosten, und jene, die wir ablehnen, haben keinen Nutzen. Wir sollten in der Lage sein, es besser zu machen.

## Kausale Situationen

Amos und ich entwarfen die Varianten des Taxi-Problems, aber den wirkmächtigen Begriff der kausalen Basisraten haben wir nicht erfunden; wir entlehnten ihn von dem Psychologen Icek Ajzen. In seinem Experiment zeigte Ajzen seinen Probanden kurze Charakterskizzen von Studenten, die in Yale eine Prüfung abgelegt hatten, und er forderte die Teilnehmer auf, die Wahrscheinlichkeit zu beurteilen, mit der jeder Student die Prüfung bestanden hatte. Die Manipulation der Basisraten war eindeutig: Ajzen sagte einer Gruppe, dass die Studenten, die ihnen vorgestellt würden, aus einem Kurs stammten, in dem 75 Prozent die Prüfung bestanden, und er sagte einer anderen Gruppe, dieselben Studenten wären Teilnehmer eines Kurses, in dem nur 25 Prozent die Prüfung bestanden hätten. Dies ist eine starke Manipulation, weil die Basisrate des Bestehens den unmittelbaren Schluss nahelegt, der Test, den nur 25 Prozent bestanden haben, müsse extrem schwierig gewesen sein. Der Schwierigkeitsgrad einer Prüfung ist natürlich einer der kausalen Faktoren, die das Ergebnis jedes Studenten bestimmen. Erwartungsgemäß wurden Ajzens Probanden in ihren Urteilen stark von den kausalen Basisraten beeinflusst, und sie gelangten zu dem Schluss, dass jeder Student in der Gruppe mit hoher Erfolgsrate mit deutlich höherer Wahrscheinlichkeit bestanden habe als ihre Kollegen in der Gruppe mit hoher Misserfolgsrate.

Ajzen benutzte eine raffinierte Methode, um eine nonkausale Basisrate zu suggerieren. Er sagte seinen Probanden, die Studenten, die ihnen kurz vorgestellt würden, stammten aus einer Stichprobe, die ihrerseits durch Selektion von Studenten, die die Prüfung bestanden oder nicht bestanden hatten, zusammengestellt worden sei. Die Information über die Gruppe mit hoher Misserfolgsrate lautete beispielsweise folgendermaßen:

> Der Forscher interessierte sich hauptsächlich für die Ursachen des Prüfungsversagens und stellte eine Stichprobe zusammen, in der 75 Prozent bei der Prüfung durchgefallen waren.

Beachten Sie den Unterschied. Diese Basisrate ist eine einfache statistische Tatsache über die Gesamtheit, aus der die Einzelfälle entnommen wurden. Sie hat keine Bedeutung für die gestellte Frage, die lautet, ob der einzelne Student die Prüfung bestanden oder nicht bestanden hat. Erwartungsgemäß hatten die ausdrücklich erwähnten Basisraten einen gewissen Einfluss auf die Urteilsbildung, aber doch in viel geringerem Umfang als die statistisch äquivalenten kausalen Basisraten. System 1 kommt gut mit Geschichten klar, in denen die Elemente kausal verknüpft sind, aber es ist schwach im statistischen Denken. Für einen bayesschen Denker sind die Versionen selbstverständlich äquivalent. Die Vermutung liegt nahe, dass wir zu einer zufriedenstellenden Schlussfolgerung gelangt sind: Kausale Basisraten werden verwendet; bloße statistische Tatsachen werden (mehr oder weniger) vernachlässigt. Die nächste Studie, eine meiner Lieblingsstudien überhaupt, zeigt, dass die Situation erheblich komplexer ist.

## Kann man Psychologie unterrichten?

Die skrupellosen Taxifahrer und die unglaublich schwierige Prüfung veranschaulichen zwei Schlüsse, die Menschen aus kausalen Basisraten ziehen. Sie sehen einerseits eine stereotype Eigenschaft, die einer Einzelperson zugeschrieben wird, und andererseits ein bedeutsames Merkmal der Situation, das sich auf die Leistung eines Individuums auswirkt. Die Teilnehmer der Experimente zogen die richtigen Schlüsse, und ihre Urteile verbesserten sich. Aber leider läuft es nicht immer so gut. Das klassische Experiment, das ich als Nächstes beschreibe, zeigt, dass Menschen aus Informationen über Basisraten keinen Schluss ziehen, der anderen Überzeugungen zuwiderläuft. Es untermauert auch die unangenehme Schlussfolgerung, dass Psychologie-Unterricht weitgehend für die Katz ist.

Das Experiment wurde vor etlichen Jahren von dem Sozialpsychologen Richard Nisbett und seinem Studenten Eugene Borgida an der Universität Michigan durchgeführt.[3] Sie erzählten Studenten von dem berühmten »Experiment zur Hilfsbereitschaft«, das ein paar Jahre zuvor an der New York University durchgeführt wurde. Die Teilnehmer dieses Experiments wurden in Einzelkabinen geführt und aufgefordert, über die Sprechanlage von ihren Privatleben und ihren Problemen zu berichten. Sie sollten nacheinander etwa zwei Minuten lang sprechen. Es war jeweils immer nur ein Mikrofon eingeschaltet. Jede Gruppe hatte sechs Probanden, von denen einer ein Strohmann war. Der Strohmann sprach als Erster und hielt sich dabei an ein Manuskript, das von den Experimentatoren vorbereitet worden war. Er erzählte, dass es ihm schwerfalle, sich in New York einzuleben, und er räumte mit unverhohlener Verlegenheit ein, dass er zu epileptischen Anfällen neige, insbesondere in Stresssituationen. Nun kamen die übrigen Teilnehmer nacheinander an die Reihe. Als das Mikrofon des Strohmanns wieder eingeschaltet wurde, wirkte er erregt und sagte, er spüre das Nahen eines Anfalls, und bat um Hilfe. Die letzten Worte, die er äußerte, waren: »Bitte … helf… Hilfe-oh-oh-oh [Würgegeräusche]. Ich … ich sterbe-er-er-er ich … sterbe-er-er … Ich … Anfall [Würgegeräusche, dann still].« In diesem Moment wurde das Mikrofon des nächsten Teilnehmers automatisch

eingeschaltet, und von der möglicherweise sterbenden Person hörte man nichts mehr.

Was, glauben Sie, haben die Versuchsteilnehmer getan? Soweit die Probanden wussten, hatte einer von ihnen einen epileptischen Anfall und um Hilfe gebeten. Allerdings gab es mehrere weitere Personen, die reagieren konnten, sodass man selbst vielleicht ruhig in seiner Kabine bleiben konnte. Dies waren die Ergebnisse: Nur vier von vierzehn Teilnehmern reagierten sofort auf den Hilferuf. Sechs haben ihre Kabine zu keinem Zeitpunkt verlassen, und fünf weitere verließen ihre Kabine erst, lange Zeit nachdem das »Anfallsopfer« scheinbar Würgegeräusche von sich gegeben hatte. Das Experiment zeigt, dass sich Menschen ihrer Verantwortung, Hilfe zu leisten, entbunden fühlen, wenn sie wissen, dass andere denselben Hilferuf gehört haben.[4]

Haben die Ergebnisse Sie überrascht? Sehr wahrscheinlich. Die meisten von uns halten sich für moralisch verantwortungsvolle Menschen, die in einer solchen Situation umgehend Hilfe leisten würden, und wir erwarten von anderen moralisch integren Menschen das Gleiche. In diesem Experiment ging es nun gerade darum, zu zeigen, dass diese Erwartung falsch ist. Selbst normale, verantwortungsbewusste Menschen eilen nicht zu Hilfe, wenn sie erwarten, dass andere die Unannehmlichkeit auf sich nehmen, sich um einen Menschen zu kümmern, der gerade einen epileptischen Anfall erleidet. Und das betrifft auch Sie.

Sind Sie bereit, folgender Aussage zuzustimmen: »Als ich die Beschreibung des Experiments über die Hilfsbereitschaft las, dachte ich, dass ich dem Fremden sofort zu Hilfe kommen würde, wie ich es sehr wahrscheinlich tun würde, wenn ich mit dem Opfer eines epileptischen Anfalls allein wäre. Vermutlich irrte ich mich. Wenn ich in eine Situation komme, in der andere Menschen die Gelegenheit haben, zu helfen, werde ich vermutlich nicht den ersten Schritt machen. Die Anwesenheit anderer würde mein persönliches Verantwortungsbewusstsein stärker verringern, als ich anfänglich dachte.« Ein Psychologie-Lehrer würde hoffen, dass Sie dies aus dem Experiment lernen. Hätten Sie selbst dieselben Schlüsse gezogen?

Der Psychologie-Professor, der das Experiment zur Hilfebereitschaft beschreibt, will, dass die Studenten die niedrige Basisrate als kausal betrachten, genau wie in dem Fall der fiktiven Yale-Prüfung. Er will, dass sie in beiden Fällen folgern, dass eine erstaunlich hohe Versagerquote auf eine sehr schwierige Prüfung hindeutet. Die Studenten sollen daraus die

Lektion lernen, dass ein bestimmendes Merkmal einer Situation, wie etwa die Verantwortungsdiffusion, normale, moralisch integre Personen – wie sie selbst – dazu veranlasst, sich erstaunlich wenig hilfsbereit zu verhalten.

Es ist sehr schwer, die eigenen Überzeugungen über die menschliche Natur zu verändern, und es ist noch schwerer, das eigene Selbstbild zu verändern. Nisbett und Borgida vermuteten, den Studenten widerstrebten die Anstrengung und die Unannehmlichkeit. Selbstverständlich wären die Studenten imstande und gewillt, bei einem Test die Einzelheiten des Hilfsbereitschafts-Experiments wiederzugeben, und sie würden sogar die »offizielle« Erklärung mit der Verantwortungsdiffusion wiederholen. Aber hatten sich ihre Anschauungen über die menschliche Natur wirklich verändert? Um dies herauszufinden, zeigten ihnen Nisbett und Borgida Videos von kurzen Interviews mit zwei Personen, die angeblich an der New Yorker Studie teilgenommen hatten. Die Interviews waren kurz und nichtssagend. Die Interviewten schienen freundliche, normale, verantwortungsvolle Menschen zu sein. Sie schilderten ihre Hobbys, ihre Freizeitaktivitäten und ihre Zukunftspläne, die ganz und gar gewöhnlich waren. Nachdem die Studenten das Video eines Interviews angesehen hatten, sollten sie einschätzen, wie schnell diese Person dem Fremden in Not zu Hilfe kommen würde.

Um die Bayessche Wahrscheinlichkeitslogik auf die Aufgabe anzuwenden, die den Studenten gestellt wurde, sollten Sie sich zuerst fragen, wie Sie die beiden Personen beurteilt hätten, wenn Sie ihre Interviews nicht gesehen hätten. Die Frage lässt sich dadurch beantworten, dass man die Basisrate heranzieht. Uns wurde mitgeteilt, dass nur vier der 15 Teilnehmer des Experiments nach der ersten Bitte Hilfe leisten wollten. Die Wahrscheinlichkeit, dass ein nicht identifizierter Teilnehmer sofort hilfsbereit ist, beträgt daher 27 Prozent. Demnach sollten Sie in Bezug auf jeden nicht näher bezeichneten Teilnehmer von der Annahme ausgehen, dass er nicht sofort Hilfe leistete. Als Nächstes verlangt die Bayessche Logik, dass Sie Ihr Urteil im Lichte relevanter Informationen über die Person anpassen. Die Videos waren jedoch gezielt so gestaltet, dass sie keine sachdienlichen Informationen enthielten; sie lieferten keine Gründe für die Annahme, dass diese Personen entweder hilfsbereiter oder weniger hilfsbereit als ein zufällig ausgewählter Student wären. Wenn keine nützlichen neuen Informationen verfügbar sind, liegt die Bayessche Lösung darin, sich an die Basisraten zu halten.

Nisbett und Borgida baten zwei Studentengruppen, sich die Videos anzusehen und das Verhalten der beiden Personen vorherzusagen. Die Studenten in der ersten Gruppe wurden nur über den Ablauf des Experiments zur Hilfsbereitschaft unterrichtet, nicht über seine Ergebnisse. In ihren Vorhersagen spiegelten sich ihre Ansichten über die menschliche Natur und ihr Verständnis der Situation wider. Wie Sie vielleicht erwarten, sagten sie vorher, dass beide Personen dem Opfer sofort zu Hilfe kommen würden. Die zweite Studentengruppe kannte sowohl den Hergang des Experiments als auch dessen Ergebnisse. Der Vergleich der Vorhersagen der beiden Gruppen liefert eine Antwort auf eine bedeutsame Frage: Lernten Studenten aus den Ergebnissen des Hilfsbereitschaftsexperiments irgendetwas, das ihre Denkweise maßgeblich veränderte? Die Antwort ist einfach: Sie lernten gar nichts. Ihre Vorhersagen über die beiden Personen unterschieden sich in keiner Weise von den Vorhersagen, die Studenten machten, die nicht über die statistischen Ergebnisse des Experiments informiert worden waren. Sie kannten die Basisrate in der Gruppe, aus der die Personen ausgewählt worden waren, aber sie blieben überzeugt davon, dass die Personen, die sie auf dem Video sahen, dem in Not geratenen Fremden umgehend zu Hilfe gekommen wären.

Für Psychologie-Lehrer sind die Folgerungen, die sich aus dieser Studie ergeben, äußerst ernüchternd. Wenn wir unseren Studenten das Verhalten der Probanden im Experiment zur Hilfsbereitschaft erläutern und es mit ihnen analysieren, erwarten wir, dass sie etwas daraus lernen; wir wollen, dass sie das Verhalten von Menschen in einer bestimmten Situation in einer neuen Weise verstehen. In der Nisbett-Borgida-Studie wurde dieses Ziel nicht erreicht, und es besteht kein Grund zu der Annahme, dass etwas anderes herausgekommen wäre, wenn sie ein anderes ungewöhnliches psychologisches Experiment ausgewählt hätten. Tatsächlich berichteten Nisbett und Borgida über ähnliche Ergebnisse in einer anderen Studie, in der leichter sozialer Druck die Probanden dazu veranlasste, viel schmerzhaftere Stromschläge zu akzeptieren, als die meisten von uns (und von ihnen) erwartet hätten. Studenten, die kein neues Verständnis für die Macht sozialer Situationen entwickeln, haben nichts Nützliches aus dem Experiment gelernt. Die Vorhersagen, die sie über zufällig ausgewählte Fremde oder über ihr eigenes Verhalten machen, deuten darauf hin, dass sie ihre Meinung darüber, wie sie sich verhalten hätten, nicht geändert haben. In den Worten von Nisbett und Borgida: »Studenten nehmen sich selbst

(und ihre Freunde und Bekannten) stillschweigend« von den Schlussfolgerungen der Experimente aus, die sie überraschen. Psychologie-Lehrer sollten allerdings nicht verzweifeln, weil Nisbett und Borgida schildern, wie es ihnen gelungen ist, ihren Studenten die Pointe des Experiments zur Hilfsbereitschaft verständlich zu machen. Sie nahmen eine neue Gruppe von Studenten und schilderten ihnen den Ablauf des Experiments, nicht aber die Gruppenergebnisse. Sie führten die beiden Videos vor und sagten ihren Studenten lediglich, die beiden Personen, die sie gerade sahen, hätten dem Fremden nicht geholfen; dann forderten sie sie auf, die Ergebnisse insgesamt abzuschätzen. Es zeigte sich, dass die Schätzungen der Studenten erstaunlich genau waren.

Um Studenten neue psychologische Kenntnisse zu vermitteln, muss man sie überraschen. Aber in welcher Weise sollte man das tun? Nisbett und Borgida fanden heraus, dass ihre Studenten, wenn sie ihnen eine überraschende statistische Tatsache mitteilten, überhaupt nichts lernten. Wenn sie die Studenten jedoch mit Einzelfällen überraschten – zwei nette Leute, die nicht geholfen hatten –, verallgemeinerten sie sofort und folgerten, dass Menschen nicht so schnell und unkompliziert helfen, wie sie geglaubt hatten. Nisbett und Borgida fassten die Ergebnisse in einem denkwürdigen Satz zusammen:

> Der Unwille der Probanden, das Besondere aus dem Allgemeinen abzuleiten, reichte allenfalls an die Bereitwilligkeit heran, mit der sie aus dem Besonderen das Allgemeine folgerten.

Dies ist ein äußerst wichtiger Befund. Menschen, denen erstaunliche statistische Tatsachen über das menschliche Verhalten beigebracht werden, mögen so beeindruckt sein, dass sie ihren Freunden davon erzählen, aber dies bedeutet nicht, dass sich ihr Verständnis der Welt wirklich verändert hat. Ob man psychologisch wirklich etwas dazugelernt hat, lässt sich nur daran ermessen, ob man ein neues Verständnis von Situationen, in die man gerät, entwickelt, nicht daran, ob man eine neue Tatsache gelernt hat. Zwischen unserem statistischen Denken und unserem Denken über Einzelfälle besteht eine gewaltige Kluft. Statistische Ergebnisse mit einer

kausalen Interpretation wirken sich stärker auf unser Denken aus als nicht kausale Informationen. Aber selbst zwingende kausale Statistiken werden langjährige Überzeugungen oder Überzeugungen, die in unseren persönlichen Erfahrungen wurzeln, nicht verändern. Andererseits haben überraschende Einzelfälle eine starke Wirkung, und sie sind ein effektiveres Instrument zur Vermittlung psychologischen Wissens, weil die Inkongruenz aufgelöst und in eine kausale Geschichte eingebaut werden muss. Aus diesem Grund enthält dieses Buch Fragen, die sich direkt an den Leser richten. Sie lernen eher etwas, wenn Sie Überraschendes in Ihrem eigenen Verhalten entdecken, als wenn Sie überraschende Tatsachen über Menschen im Allgemeinen hören.

## Zum Thema »Ursachen und Statistiken«

> »Wir können nicht davon ausgehen, dass sie aus rein statistischen Darlegungen etwas lernen werden. Wir sollten ihnen deshalb einen oder zwei repräsentative Einzelfälle zeigen, um ihr System 1 zu beeinflussen.«

> »Es gibt keinen Grund, zu befürchten, dass diese statistische Information ignoriert wird. Im Gegenteil, sie wird sofort dazu benutzt werden, ein Stereotyp zu bekräftigen.«

# *17. Regression zum Mittelwert*

Eines der befriedigendsten Heureka-Erlebnisse meiner beruflichen Laufbahn hatte ich, als ich Fluglehrer der israelischen Luftwaffe in der Psychologie erfolgreichen Trainings unterrichtete. Ich erzählte ihnen von einem wichtigen Grundsatz erfolgreicher Schulung von Fähigkeiten: Die Belohnung von Leistungssteigerungen ist effektiver als die Bestrafung von Fehlern. Zahlreiche Ergebnisse von Studien an Tauben, Ratten, Menschen und anderen Säugetieren erhärten diese These.

Als ich meinen begeisterten Vortrag beendet hatte, hob einer der erfahrensten Ausbilder in der Gruppe die Hand und hielt selbst eine kurze Rede. Zunächst räumte er ein, dass die Belohnung von Leistungsverbesserungen bei Vögeln eine gute Idee sein mag, aber er meinte, dass es mit Sicherheit für Flugkadetten nicht optimal wäre. Er sagte: »Bei vielen Gelegenheiten habe ich Flugkadetten für die saubere Ausführung bestimmter Flugmanöver gelobt. Wenn sie dann das gleiche Manöver wieder probieren, sind sie meistens schlechter. Andererseits habe ich oft in den Kopfhörer eines Kadetten gebrüllt, weil ich mit seiner Leistung nicht zufrieden war, und beim nächsten Versuch hat er es dann im Allgemeinen besser gemacht. Also erzählen Sie uns bitte nicht, dass Belohnungen immer den gewünschten Erfolg bringen und Bestrafungen nicht, denn das Gegenteil ist der Fall.«

Dies war ein freudiger Moment der Erkenntnis, als ich ein Prinzip der Statistik, das ich seit Jahren unterrichtet hatte, in einem neuen Licht sah. Der Ausbilder hatte recht – aber zugleich hatte er vollkommen unrecht! Seine Bemerkung war scharfsinnig und richtig: Wenn er einmal eine Leistung lobte, folgte mit hoher Wahrscheinlichkeit beim nächsten Mal eine enttäuschende Leistung, und Tadel führte in der Regel zu einer Verbesserung der Leistung. Aber der Schluss, den er aus seinen Erfahrungen über die Effizienz von Belohnung und Bestrafung zog, war völlig verfehlt. Was er beobachtet hatte, war die sogenannte »Regression zum Mittelwert«, die in diesem Fall auf zufällige Schwankungen in der Qualität der Leistung zurückzuführen war. Verständlicherweise lobte er einen Kadetten nur dann, wenn dieser eine weit überdurchschnittliche

Leistung brachte. Aber der Kadett hatte bei der konkreten Übung vermutlich einfach nur Glück, und seine Leistung würde sich beim nächsten Mal wahrscheinlich verschlechtern, unabhängig davon, ob er gelobt wurde oder nicht. In ähnlicher Weise würde der Ausbilder den Kadetten nur dann über Funk anherrschen, wenn dieser eine ungewöhnlich schlechte Leistung zeigte, die sich unabhängig von dem, was der Ausbilder tat, wahrscheinlich verbessern würde. Der Ausbilder hatte an die zwangsläufigen Schwankungen eines zufallsabhängigen Prozesses eine kausale Interpretation angefügt.

Die Kritik verlangte nach einer Erwiderung, aber eine Lektion in der Algebra von Vorhersagen würde nicht begeistert aufgenommen werden. Stattdessen zeichnete ich mit Kreide ein Ziel auf den Boden. Ich bat jeden anwesenden Offizier, sich mit dem Rücken zum Ziel aufzustellen und direkt hintereinander zwei Münzen daraufzuwerfen, ohne sich danach umzusehen. Wir maßen die Entfernungen vom Ziel und schrieben die beiden Ergebnisse jedes Teilnehmers an die Tafel. Dann schrieben wir eine weitere Liste, in der wir die Ergebnisse nach der Treffgenauigkeit ordneten – von der besten zur schlechtesten Leistung beim ersten Versuch. Es war offensichtlich, dass die meisten (aber nicht alle) derjenigen, die beim ersten Mal am besten abgeschnitten hatten, beim zweiten Versuch schlechtere Ergebnisse erzielten und dass diejenigen, die beim ersten Versuch eine schwache Leistung zeigten, sich im Allgemeinen verbesserten. Ich wies die Ausbilder darauf hin, dass das, was Sie an der Tafel sahen, mit dem übereinstimmte, was wir über die Leistung bei den Flugmanövern bei aufeinanderfolgenden Trainingsdurchläufen gehört hatten: Auf eine schwache Leistung folgte in der Regel eine Verbesserung und auf eine gute Leistung eine Verschlechterung, ohne dass dabei Lob oder Tadel die geringste Rolle gespielt hätten.

An jenem Tag wurde mir klar, dass die Fluglehrer sich von einer fatalen Kontingenz täuschen ließen: Weil sie Kadetten rügten, wenn diese eine schlechte Leistung erbrachten, wurden sie zumeist mit einer anschließenden Verbesserung belohnt, selbst wenn der Tadel faktisch völlig wirkungslos blieb. Aber diese leidige kognitive Verzerrung betrifft nicht nur die Ausbilder. Ich war zufällig auf eine weitreichende Tatsache der menschlichen Existenz gestoßen: Das Alltagsleben liefert uns verzerrte Feedbacks. Weil wir dazu tendieren, Menschen, die uns gefällig sind, freundlich zu behandeln, während wir zu denen, die es nicht sind, gemein

sind, werden wir aus statistischen Gründen dafür bestraft, nett zu sein, und dafür belohnt, gehässig zu sein.

## Talent und Glück

Vor ein paar Jahren bat John Brockman, der Herausgeber des Online-Magazins *Edge*, eine Reihe von Wissenschaftlern, ihre »Lieblingsgleichung« zu nennen. Ich nannte ihm die folgenden:

Erfolg = Talent + Glück

Großer Erfolg = ein wenig mehr Talent + viel Glück

Die wenig überraschende Ansicht, dass Glück oftmals zu Erfolg beiträgt, hat überraschende Konsequenzen, wenn wir sie auf die ersten beiden Tage eines hochkarätigen Golfturniers anwenden. Der Einfachheit halber wollen wir annehmen, dass an beiden Tagen der durchschnittliche Score der Wettbewerber bei einem Par von 72 lag. Wir konzentrieren uns auf einen Spieler, der am ersten Tag sehr gut spielte und mit einem Score von 66 abschloss. Was sagt uns dieser hervorragende Score? Die Schlussfolgerung drängt sich auf, dass der Golfspieler talentierter ist als der durchschnittliche Turnierteilnehmer. Die Erfolgsformel deutet darauf hin, dass ein anderer Schluss genauso gerechtfertigt ist: Der Golfer, der am ersten Tag so gut spielte, hatte an diesem Tag einfach überdurchschnittlich viel Glück. Wenn man annimmt, dass sowohl Begabung als auch Glück zum Erfolg beitragen, ist die Schlussfolgerung, dass der erfolgreiche Golfer Glück hatte, genauso berechtigt wie die, dass er begabt ist. Aus dem gleichen Grund haben wir bei einem Spieler, der am selben Tag einen Score von fünf über Par spielte, Grund zu der Schlussfolgerung, dass er eher schwach ist *und* einen schlechten Tag hatte. Wir wissen natürlich, dass keiner dieser beiden Schlüsse zweifellos zutreffend ist. Es ist durchaus möglich, dass der Spieler mit einem Score von 77 sehr talentiert ist, aber einen ungewöhnlich schlechten Tag hatte. Doch die folgenden Schlüsse aus dem Score von Tag

eins sind plausibel, auch wenn sie nicht sicher sind, und sie sind häufiger richtig als falsch.

Überdurchschnittlicher Score an Tag eins
= überdurchschnittliches Talent + Glück an Tag eins

Und:

Unterdurchschnittlicher Score an Tag eins
= unterdurchschnittliches Talent + Pech an Tag eins

Nehmen wir jetzt an, Sie kennen den Score eines Golfers an Tag eins und sollen seinen Score an Tag zwei vorhersagen. Sie erwarten, dass der Golfer am zweiten Tag nichts von seiner Begabung eingebüßt hat, sodass Sie für den ersten Spieler eine überdurchschnittliche Leistung vorhersagen und für den zweiten Spieler eine unterdurchschnittliche. Mit dem Glück verhält es sich allerdings anders. Da Sie das Glück des Golfers am zweiten (oder einem anderen Tag) nicht vorhersagen können, müssen Sie annehmen, dass es durchschnittlich sein wird – weder besonders positiv noch besonders negativ. Dies bedeutet, dass Sie bei der Einschätzung der Scores der Spieler an Tag zwei nicht einfach davon ausgehen sollten, dass sie ihre Leistung von Tag eins wiederholen werden, sofern Sie über keine zusätzlichen Informationen verfügen. Sie können bestenfalls folgende Aussagen treffen:

– Der Golfer, der an Tag eins eine gute Leistung zeigte, wird vermutlich auch an Tag zwei erfolgreich sein, aber weniger erfolgreich als am ersten Tag, weil das außerordentliche Glück, das er vermutlich an Tag eins hatte, voraussichtlich nicht anhalten wird.
– Der Golfer, der an Tag eins eine schlechte Leistung brachte, wird an Tag zwei vermutlich ebenfalls eine unterdurchschnittliche Leistung zeigen, aber er wird sich verbessern, weil seine wahrscheinliche Pechsträhne vermutlich nicht andauern wird.

Wir erwarten auch, dass sich der Unterschied zwischen den beiden Golfern am zweiten Tag verringert, auch wenn der erste Spieler unserer Einschätzung nach auch am zweiten Tag eine bessere Leistung erbringen wird.

Meine Studenten waren immer überrascht zu hören, dass die beste vorhergesagte Leistung für Tag zwei mittelmäßiger ist, näher am Durchschnitt liegt als die Daten, auf die sich die Vorhersage stützt (der Score an Tag eins). Aus diesem Grund wird dieses Muster Regression zum Mittelwert genannt. Je extremer der ursprüngliche Score ist, umso mehr Regression erwarten wir, weil ein außerordentlich guter Score auf einen Tag hindeutet, an dem der Spieler sehr viel Glück hatte. Die regressive Vorhersage ist vernünftig, aber es ist nicht sicher, ob sie zutreffen wird. Einige wenige der Golfer, die an Tag eins einen Score von 66 erzielten, werden am zweiten Tag ein noch besseres Ergebnis erreichen, wenn sie noch mehr Glück haben. Die meisten werden schlechtere Leistungen bringen, weil sie kein überdurchschnittliches Glück mehr haben werden.

Lassen Sie uns jetzt eine Betrachtung gegen den Zeitpfeil vornehmen. Wir stellen eine Rangfolge der Spieler nach ihrer Leistung an Tag zwei auf und betrachten dann ihre Leistung an Tag eins. Wir stoßen auf genau das gleiche Muster der Regression zum Mittelwert. Die Golfer, die an Tag zwei die besten Ergebnisse erzielten, hatten vermutlich an diesem Tag Glück, und die beste Schätzung ist, dass sie an Tag eins nicht so viel Glück hatten und weniger gute Ergebnisse erzielten. Die Tatsache, dass wir eine Regression beobachten, wenn wir ein früheres Ereignis auf der Basis eines späteren Ereignisses vorhersagen, sollte Sie davon überzeugen, dass diese Regression nicht kausal erklärt werden kann.

Regressionseffekte sind allgegenwärtig, und das Gleiche gilt für irrige kausale Geschichten, die sie erklären sollen. Ein bekanntes Beispiel ist der »*Sports Illustrated*-Pechvogel«, die Behauptung, dass ein Sportler, dessen Bild auf der Titelseite des Magazins erscheint, in der nächsten Saison zwangsläufig eine schlechte Leistung bringen wird. Überzogenes Selbstvertrauen und hoher Erwartungsdruck werden oft als Erklärungen angeführt. Aber es gibt eine einfachere Erklärung für den Pechvogel: Ein Sportler, der es auf die Titelseite von *Sports Illustrated* schafft, muss in der vorangegangenen Saison eine außerordentlich gute Leistung gezeigt haben, und dabei hat höchstwahrscheinlich das Glück nachgeholfen – aber das Glück ist unbeständig.

Ich habe zufälligerweise im Fernsehen das Skispringen der Männer bei der Winterolympiade angesehen, während Amos und ich einen Artikel über intuitive Vorhersagen schrieben. In diesem Wettbewerb hat jeder Sportler zwei Sprünge, und die Addition der Ergebnisse beider Durchgänge ergibt den Endstand. Mit Erstaunen hörte ich die Kommentare des Sportreporters, während sich die Sportler auf den zweiten Sprung vorbereiteten: »Norwegen hatte einen hervorragenden ersten Sprung; er wird angespannt sein, alles daransetzen, seine Führung zu behaupten, und vermutlich nicht mehr so weit springen.« Oder: »Schweden hatte einen schlechten ersten Sprung, und er weiß jetzt, dass er nichts zu verlieren hat, und wird daher entspannt sein, was ihm helfen sollte, eine bessere Leistung zu bringen.« Der Kommentator hatte offensichtlich die Regression zum Mittelwert bemerkt und eine kausale Geschichte erfunden, für die es keine Anhaltspunkte gab. Die Geschichte selbst konnte sogar wahr sein. Wenn wir den Puls der Sportler vor jedem Sprung messen würden, würden wir vielleicht feststellen, dass sie nach einem schlechten ersten Sprung tatsächlich entspannter sind. Vielleicht aber auch nicht. Wir sollten uns merken, dass der Unterschied zwischen dem ersten und dem zweiten Sprung keine kausale Erklärung benötigt. Er ist eine mathematisch unvermeidliche Folge der Tatsache, dass Glück bei dem Ergebnis des ersten Sprungs eine Rolle spielte. Kein sehr befriedigendes Fazit – uns allen wäre eine kausale Erklärung lieber –, aber mehr hat es damit nicht auf sich.

## Regression verstehen

Das Phänomen der Regression ist, ganz gleich, ob es unbemerkt bleibt oder falsch erklärt wird, dem menschlichen Geist fremd. Tatsächlich so fremd, dass es erst 200 Jahre nach der Erfindung der Gravitationstheorie und der Differenzialrechnung erstmals entdeckt und verstanden wurde. Außerdem bedurfte es eines der klügsten Köpfe Großbritanniens im 19. Jahrhundert, um das Phänomen zu verstehen, und dies gelang ihm nur mit großer Mühe.

Sir Francis Galton, ein Halbcousin von Charles Darwin und ein berühmter Universalgelehrter, entdeckte die Regression zum Mittelwert und gab ihr ihren Namen. Man spürt den Nervenkitzel des Entdeckers in einem Artikel, den er 1886 unter dem Titel »Regression towards Mediocrity in Hereditary Stature« (»Regression zum Mittelmaß beim hereditären Wuchs«) veröffentlichte, wo er über Größenmessungen an aufeinanderfolgenden Generationen von Samen und über Vergleiche zwischen der Körpergröße von Kindern und der ihrer Eltern berichtete. Über seine Studien an Samen schreibt er:

> Sie erbrachten Ergebnisse, die bemerkenswert zu sein schienen, und ich benutzte sie als Grundlage einer Vorlesung an der Royal Institution am 9. Februar 1877. Diese Experimente ergaben, dass die Nachkommen hinsichtlich der Größe ihren Elternsamen im Allgemeinen *nicht* ähnelten, sondern immer zum Mittelmaß tendierten – kleiner als die Eltern waren, wenn die Eltern groß waren; größer als die Eltern waren, wenn die Eltern sehr klein waren … Die Experimente zeigten des Weiteren, dass die mittlere Regression der Filialgeneration zum Mittelmaß direkt proportional zur Abweichung der Parentalgeneration vom Mittelmaß gewesen ist.

Galton erwartete offensichtlich, dass seine gelehrten Zuhörer bei der Royal Institution – der ältesten unabhängigen Forschungsgesellschaft der Welt – von seiner »bemerkenswerten Beobachtung« genauso überrascht wären,

wie er es gewesen war. Das eigentlich Bemerkenswerte ist, dass er von einer statistischen Regelmäßigkeit überrascht war, die genauso weit verbreitet ist wie die Luft, die wir atmen. Regressionseffekte sind allgegenwärtig, aber wir erkennen sie nicht als solche. Obwohl für jedermann sichtbar, bleiben sie unerkannt. Galton brauchte Jahre, um von seiner Entdeckung der Größenregression in der Tochtergeneration zu dem allgemeinen Konzept zu gelangen, dass es zwangsläufig zu einer Regression kommt, wenn die Korrelation zwischen zwei Messgrößen nicht perfekt ist, und er brauchte die Hilfe der hervorragendsten Statistiker seiner Zeit, um zu dieser Schlussfolgerung zu gelangen.[1]

Eine der Hürden, die Galton überwinden musste, war das Problem der Messung der Regression zwischen Variablen, die auf verschiedenen Skalen gemessen werden, wie etwa das Körpergewicht und die Fähigkeit, Klavier zu spielen. Dies erreicht man dadurch, dass man die statistische Gesamtheit als Standardbezugsgröße verwendet. Nehmen wir an, das Gewicht und die Fähigkeit, Klavier zu spielen, seien für hundert Kinder in allen Klassenstufen einer Grundschule gemessen worden, und in beiden Variablen seien sie von hoch nach niedrig eingestuft worden. Wenn Jane beim Klavierspielen an dritter Stelle steht und beim Gewicht an 27. Stelle, ist es sachgerecht, zu sagen, dass sie im Klavierspiel-Ranking viel weiter vorn liegt als im Gewichts-Ranking. Machen wir nun einige Annahmen, die die Dinge erleichtern werden:

In jedem Alter

- hängt die Leistungsfähigkeit beim Klavierspiel nur von der Zahl der wöchentlichen Übungsstunden ab.
- hängt das Gewicht nur vom Verzehr von Eiscreme ab.
- besteht kein Zusammenhang zwischen dem Eiscremeverzehr und der Zahl der wöchentlichen Übungsstunden.

Jetzt können wir unter Verwendung von Rangzahlen (oder »Standardwerten«,[2] wie sie von den Statistikern vorgezogen werden) einige Gleichungen schreiben:

Gewicht = Alter + Eiscremeverzehr

Klavierspiel = Alter + Anzahl der wöchentlichen Übungsstunden

Sie können erkennen, dass es zu einer Regression zum Mittelwert kommt, wenn wir das Leistungsniveau beim Klavierspielen aus dem Gewicht vorhersagen oder umgekehrt. Wenn wir über Tom lediglich wissen, dass er den zwölften Rang im Gewichts-Ranking (weit über dem Durchschnitt) einnimmt, kann man (statistisch) daraus folgern, dass er vermutlich überdurchschnittlich alt ist und mehr Eiscreme verzehrt als andere Kinder. Wenn wir über Barbara nur wissen, dass sie an 85. Stelle im Klavierspiel-Ranking steht (weit unterhalb des Gruppendurchschnitts), können wir daraus folgern, dass sie vermutlich jung ist und weniger übt als die meisten anderen Kinder.

Der Korrelationskoeffizient zwischen zwei Messwerten, der sich zwischen null und eins bewegt, ist ein Maß des relativen Gewichts der Faktoren, die ihnen gemeinsam sind. So haben wir beispielsweise die Hälfte unserer Gene mit jedem unserer Elternteile gemeinsam, und bei Merkmalen, auf die Umweltfaktoren einen vergleichsweise geringen Einfluss haben, wie etwa die Körpergröße, liegt die Korrelation zwischen Eltern und Kind bei 0,50.[3] Nachfolgend zum Verständnis des Korrelationsmaßes einige Beispiele von Koeffizienten:

- Die Korrelation zwischen den Größen von Objekten, die präzise in englischen oder in metrischen Maßeinheiten gemessen werden, ist eins. Jeder Faktor, der ein Maß beeinflusst, beeinflusst auch das andere; 100 Prozent der Determinanten sind gemeinsam.
- Die Korrelation zwischen der selbst angegebenen Körpergröße und dem Gewicht bei erwachsenen amerikanischen Männern beträgt 0,41.[4] Wenn man Frauen und Kinder einbezieht, ist die Korrelation viel höher, weil die Geschlechtszugehörigkeit und das Alter der Individuen ihre Größe und ihr Gewicht beeinflussen, was das relative Gewicht der gemeinsamen Faktoren steigert.
- Die Korrelation zwischen den Punktwerten im Studierfähigkeitstest und dem Notendurchschnitt im College-Abschlusszeugnis beträgt ungefähr 0,60. Die Korrelation zwischen Studierfähigkeitstests und dem Studienerfolg ist viel geringer, vor allem deshalb, weil die

gemessene Fähigkeit in dieser ausgewählten Gruppe kaum variiert. Wenn jeder eine ähnliche Befähigung besitzt, spielen Unterschiede in dieser Maßgröße vermutlich keine große Rolle bei Erfolgsmaßstäben.

- Die Korrelation zwischen Einkommen und Bildungsniveau in den Vereinigten Staaten beträgt ungefähr 0,40.[5]
- Die Korrelation zwischen dem Einkommen einer Familie und den letzten vier Ziffern ihrer Telefonnummer ist null.

Francis Galton brauchte mehrere Jahre, um herauszufinden, dass Korrelation und Regression nicht zwei getrennte Konzepte sind – vielmehr sind es unterschiedliche Perspektiven auf das gleiche Konzept.[6] Die allgemeine Regel ist einfach, hat jedoch überraschende Folgen: Immer wenn zwei Messwerte nicht perfekt miteinander korreliert sind, kommt es zu einer Regression zum Mittelwert. Um Galtons Erkenntnis zu verdeutlichen, wollen wir folgende Aussage betrachten, die die meisten Menschen ziemlich interessant finden:

> Hochintelligente Frauen heiraten tendenziell Männer, die weniger intelligent sind als sie.

Sie können auf einer Party eine nette Unterhaltung beginnen, indem Sie nach einer Erklärung fragen, und Ihre Freunde werden Ihrer Bitte bereitwillig nachkommen. Selbst Personen, die gewisse statistische Kenntnisse besitzen, werden die Aussage spontan kausal interpretieren. Manche werden die Meinung äußern, hochintelligente Frauen wollten der Rivalität durch ebenso intelligente Männer aus dem Weg gehen, oder sie seien bei ihrer Partnerwahl zu Kompromissen gezwungen, weil intelligente Männer nicht mit intelligenten Frauen konkurrieren wollten. Bei einer guten Party werden auch weiter hergeholte Erklärungen aufgetischt werden. Betrachten wir nunmehr folgende Aussage:

> Die Korrelation zwischen den IQ-Werten von Ehegatten ist

weniger als perfekt.

Die Aussage ist offenkundig wahr und völlig uninteressant. Wer würde eine perfekte Korrelation erwarten? Es gibt nichts zu erklären. Aber die Aussage, die Sie interessant gefunden haben, und die Aussage, die Sie trivial gefunden haben, sind algebraisch äquivalent. Wenn die Korrelation zwischen der Intelligenz von Ehegatten weniger als perfekt ist (und wenn Männer und Frauen im Durchschnitt nicht unterschiedlich intelligent sind), dann ist es mathematisch unausweichlich, dass hochintelligente Frauen mit Männern verheiratet sind, die im Schnitt weniger intelligent sind als sie (und umgekehrt natürlich). Die beobachtete Regression zum Mittelwert kann nicht interessanter oder erklärbarer sein als die imperfekte Korrelation.

Vermutlich haben Sie Verständnis für Galtons Ringen mit dem Regressionsbegriff. Tatsächlich pflegte der Statistiker David Freedman zu sagen, dass, wenn in einem straf- oder zivilrechtlichen Verfahren das Thema Regression aufkommt, jene Seite, die die Regression erklären muss, den Prozess verlieren wird. Weshalb ist dieses Phänomen so schwer zu verstehen? Der Hauptgrund für die Schwierigkeit ist ein wiederkehrendes Thema dieses Buches: Unser Intellekt zeichnet sich durch eine starke Neigung zu kausalen Erklärungen aus und kommt nicht gut mit »bloßer Statistik« zurecht. Wenn unsere Aufmerksamkeit auf ein Ereignis gelenkt wird, sucht das assoziative Gedächtnis nach dessen Ursache – genauer gesagt, breitet sich die Aktivierung automatisch auf jede Ursache aus, die bereits im Gedächtnis gespeichert ist. Kausale Erklärungen werden abgerufen, wenn eine Regression erkannt wird, aber sie sind falsch, weil Regressionen zum Mittelwert eine Erklärung, aber keine Ursache haben. Das Ereignis, das bei dem Golfturnier unsere Aufmerksamkeit erregt, ist die häufige Leistungsverschlechterung bei den Golfern, die an Tag eins erfolgreich waren. Die lässt sich am besten damit erklären, dass diese Golfer an jenem Tag ungewöhnlich viel Glück hatten, aber dieser Erklärung fehlt die kausale Kraft, die unser Intellekt bevorzugt. Tatsächlich bezahlen wir bestimmte Personen sehr gut dafür, dass sie interessante Erklärungen für Regressionseffekte liefern. Ein Wirtschaftskommentator, der richtigerweise feststellt, dass »das Unternehmen dieses Jahr bessere

Ergebnisse erzielte, weil seine Geschäfte letztes Jahr so schlecht liefen«, wird vermutlich nicht lange auf Sendung sein.

Unsere Schwierigkeiten mit dem Regressionsbegriff rühren sowohl von System 1 als auch von System 2 her. Ohne spezielle Schulung und in nicht wenigen Fällen sogar nach einer statistischen Unterweisung bleibt der Zusammenhang zwischen Korrelation und Regression im Dunkeln. System 2 fällt es schwer, zu verstehen und zu lernen. Dies hängt zum Teil mit der beharrlichen Forderung nach kausalen Interpretationen zusammen, die ein Merkmal von System 1 ist.

> Bei depressiven Kindern, die mit einem Energiedrink behandelt werden, verbessert sich die Stimmungslage deutlich über einen dreimonatigen Zeitraum.

Ich habe diese Schlagzeile erfunden, aber die Tatsache, über die dort berichtet wird, ist wahr: Wenn man eine Gruppe depressiv verstimmter Kinder eine Zeit lang mit einem Energiedrink behandeln würde, würden sie eine klinisch signifikante Verbesserung zeigen. Es ist ebenso der Fall, dass sich bei depressiven Kindern, die jeden Tag eine Zeit lang Kopfstand machen oder zwanzig Minuten eine Katze streicheln, die Gemütslage aufhellt. Die meisten Leser solcher Schlagzeilen werden unwillkürlich folgern, dass der Energiedrink oder das Streicheln der Katze die Stimmungsaufhellung verursachte, aber diese Schlussfolgerung ist völlig ungerechtfertigt. Depressive Kinder sind eine Extremgruppe, sie sind bedrückter als die meisten anderen Kinder – und Extremgruppen regredieren mit der Zeit zum Mittelwert. Die Korrelation zwischen Depressionsscores in aufeinanderfolgenden Depressionstests ist nicht perfekt, sodass es zu einer Regression zum Mittelwert kommt: Die Stimmungslage depressiver Kinder wird sich mit der Zeit auch dann in einem gewissen Umfang verbessern, wenn sie keine Katzen streicheln und keinen Energiedrink zu sich nehmen. Um den Schluss zu ziehen, dass ein besonderes Getränk – oder eine andere Therapie – wirksam ist, müssen Sie eine Gruppe von Patienten, die diese Behandlung erhält, mit einer Kontrollgruppe vergleichen, die nicht behandelt wird (oder, besser noch, ein Placebo bekommt). Von der Kontrollgruppe erwartet man, dass sich ihr

Zustand allein durch Regression verbessert, und das Experiment dient dem Zweck, zu klären, ob sich der Zustand der behandelten Patienten über das Maß, das sich mit der Regression erklären lässt, hinaus verbessert hat.

Falsche kausale Interpretationen von Regressionseffekten beschränken sich nicht auf die Leser der Boulevardpresse. Der Statistiker Howard Wainer hat eine lange Liste bedeutender Forscher zusammengestellt, die denselben Fehler gemacht haben – bloße Korrelation mit Kausalität zu verwechseln.[7] Regressionseffekte sind eine häufige Quelle von falschen Schlussfolgerungen in wissenschaftlichen Studien, und erfahrene Wissenschaftler entwickeln eine gesunde Furcht vor der Falle ungerechtfertigter kausaler Schlüsse.

Eines meiner Lieblingsbeispiele für die Fehler, die sich bei intuitiven Vorhersagen einschleichen, ist dem hervorragenden Buch *Judgement in Managerial Decision Making* von Max Bazerman entnommen:

> Sie erstellen Umsatzprognosen für eine Warenhauskette. Alle Geschäfte sind ungefähr gleich groß und haben ein ähnliches Warensortiment, aber ihre Umsätze unterscheiden sich wegen des Standorts, der Konkurrenzlage und zufälliger Faktoren. Sie erhalten die Zahlen für das Jahr 2011 und sollen den Umsatz für 2012 vorhersagen. Sie sollen die prognostizierte gesamtwirtschaftliche Wachstumsquote von 10 Prozent berücksichtigen. Wie würden Sie die nachfolgende Tabelle vervollständigen?

| **Geschäft** | **2011** | **2012** |
|---|---|---|
| 1 | 11 Mio. Dollar | _____ |
| 2 | 23 Mio. Dollar | _____ |

| | | |
|---|---|---|
| 3 | 18 Mio. Dollar | _____ |
| 4 | 29 Mio. Dollar | _____ |
| Summe | 81 Mio. Dollar | 89,1 Mio. Dollar |

**Tabelle 2**

Nachdem Sie dieses Kapitel gelesen haben, wissen Sie, dass die naheliegende Lösung – 10 Prozent zum Umsatz jedes Geschäfts hinzuaddieren – falsch ist. Sie wollen, dass Ihre Prognosen regressiv sind, und dazu ist es notwendig, dass Sie bei den schwachen Niederlassungen mehr als 10 Prozent addieren und bei den anderen weniger (oder sogar etwas vom Vorjahresumsatz subtrahieren). Aber wenn Sie andere Menschen fragen, werden Sie wahrscheinlich für Verwirrung sorgen: Weshalb behelligen Sie mich mit einer Frage, deren Antwort auf der Hand liegt? Wie Galton schmerzlich erfuhr, ist das Konzept der Regression alles andere als selbstverständlich.

## Zum Thema »Regression zum Mittelwert«

»Sie sagt, die Erfahrung habe sie gelehrt, dass Kritik wirksamer sei als Lob. Sie versteht nicht, dass all dies auf die Regression zum Mittelwert zurückzuführen ist.«

»Vielleicht war sein zweites Vorstellungsgespräch weniger eindrucksvoll als das erste, weil er Angst hatte, uns zu enttäuschen, aber es ist wahrscheinlicher, dass sein erstes außergewöhnlich gut war.«

»Unser Ausleseverfahren ist gut, aber nicht perfekt, sodass wir mit einer Regression rechnen. Es sollte uns nicht überraschen, wenn die besten Kandidaten unsere Erwartungen nicht erfüllen.«

# *18. Intuitive Vorhersagen bändigen*

Das Leben bietet uns viele Gelegenheiten zur Vorhersage. Ökonomen prognostizieren die Inflationsrate und die Arbeitslosigkeit, Finanzanalysten prognostizieren den Gewinn von Unternehmen, Militärexperten prognostizieren Verluste, Wagniskapitalgesellschaften bewerten die Rentabilität möglicher Investments, Bauunternehmen schätzen den Zeitbedarf für ihre Projekte ab, Küchenchefs antizipieren die Nachfrage nach den Gerichten auf ihren Speisekarten, Ingenieure schätzen die Menge an Beton, die sie für ein Gebäude benötigen, die Einsatzleiter der Feuerwehr schätzen die Anzahl der Löschfahrzeuge ab, die sie zur Bekämpfung eines Brandes benötigen. In unserem Privatleben sagen wir die Reaktion unseres Partners auf eine bestimmte Verhaltensweise oder Äußerung vorher oder unsere zukünftige Anpassung an einen neuen Arbeitsplatz.

Einige prognostische Urteile, wie sie etwa von Ingenieuren getroffen werden, stützen sich weitgehend auf sogenannte Lookup-Tabellen, präzise Berechnungen und explizite Analysen von Ergebnissen, die bei ähnlichen Gelegenheiten beobachtet wurden. Andere beruhen auf Intuitionen und beziehen das System 1 ein; dabei lassen sich zwei Grundtypen von Intuitionen unterscheiden. Einige Intuitionen basieren hauptsächlich auf Fertigkeit und Sachkunde, die durch regelmäßige Übung erworben werden. Die schnellen und automatischen Urteile und Entscheidungen von Schachmeistern, Einsatzleitern der Feuerwehr und Ärzten, die Gary Klein in *Natürliche Entscheidungsprozesse: Über die »Quellen der Macht«, die unsere Entscheidungen lenken* und anderen Büchern beschrieben hat, verdeutlichen diese fachkundigen Intuitionen, bei denen Experten schnell eine Lösung für das aktuelle Problem einfällt, weil vertraute Hinweisreize erkannt werden.

Andere Intuitionen, die manchmal subjektiv nicht von den ersten zu unterscheiden sind, gehen aus Heuristiken hervor, die oftmals eine schwierige Frage durch eine leichtere ersetzen. Intuitive Urteile lassen sich auch dann mit einem hohen Grad subjektiver Überzeugung treffen, wenn sie auf nicht regressiven Bewertungen schwacher Informationen basieren.

Selbstverständlich werden viele Urteile, insbesondere im professionellen Bereich, durch eine Kombination von Analyse und Intuition beeinflusst.

## Nicht regressive Intuitionen

Kehren wir zurück zu einer Person, die wir bereits kennengelernt haben:

> Julie studiert im vierten Jahr an einer staatlichen Universität. Im Alter von vier Jahren hat sie flüssig gelesen. Was ist ihr Notendurchschnitt?

Menschen, die sich im amerikanischen Bildungssystem auskennen, warten schnell mit einer Zahl auf, die oftmals im Bereich von 3,7 bis 3,8 liegt. Wie kommt es dazu?

Daran sind mehrere Operationen von System 1 beteiligt.

– Es wird nach einer kausalen Verknüpfung zwischen der Information ( Julies Lesefähigkeit) und dem Ziel der Vorhersage (ihrem Notendurchschnitt) gesucht. Es kann auch eine indirekte Verknüpfung sein. In diesem Beispiel sind sowohl frühe Lesefähigkeit als auch die Tatsache, dass sie in einem Wohlstandsland mit hohem Bruttoinlandsprodukt studiert, Hinweise auf gute Studienleistungen. Irgendein Zusammenhang ist notwendig. Sie (Ihr System 2) würden vermutlich einen Bericht darüber, dass Julie einen Wettbewerb im Fliegenfischen gewonnen hat oder dass sie sich auf der Highschool beim Gewichtheben hervorgetan hat, als irrelevant abtun. Tatsächlich handelt es sich um einen dichotomen Prozess. Wir können Informationen als irrelevant oder falsch verwerfen, aber System 1 ist nicht in der Lage, kleinere Unzulänglichkeiten in den Informationen zu berücksichtigen. Infolgedessen sind intuitive Vorhersagen völlig unempfindlich gegenüber der tatsächlichen prognostischen Qualität der Informationen. Wenn eine Verknüpfung gefunden wird, wie im Fall von Julies früher Lesefähigkeit, gilt die WYSIATI-Regel: »Nur was man (gerade) weiß, zählt.« Ihr assoziatives Gedächtnis konstruiert aus den verfügbaren Informationen die bestmögliche Geschichte.

– Als Nächstes werden die Informationen in Bezug auf eine relevante Norm beurteilt. Wie frühreif ist ein Kind, das mit vier Jahren flüssig liest? Welchem relativen Rang beziehungsweise Perzentilwert (Prozentrang) entspricht diese Leistung? Die Gruppe, mit der das Kind verglichen wird (seine Vergleichsgruppe), ist nicht näher beschrieben, aber dies ist auch in der alltäglichen Kommunikation der Fall: Wenn jemand, der seinen College-Abschluss macht, als »recht intelligent« beschrieben wird, muss man nur selten fragen: »Wenn Sie ›recht intelligent‹ sagen, an welche Vergleichsgruppe denken Sie dann?«
– Der nächste Schritt umfasst Ersetzung und Intensitätsabstimmung. Die Bewertung der schwachen Informationen über die kognitiven Fähigkeiten im Kindesalter tritt an die Stelle der Antwort auf die Frage nach Julies Notendurchschnitt im College. Julie wird für ihren Notendurchschnitt und ihre frühkindliche Leseleistung der gleiche Perzentilwert zugeschrieben.
– Die Frage gab vor, dass die Antwort auf der Notendurchschnittsskala liegen muss, was eine weitere Intensitätsabstimmung erfordert, zwischen einem allgemeinen Eindruck von Julies Studienleistungen und jenem Notendurchschnitt, der den Informationen über ihre Begabung entspricht. Der letzte Schritt besteht darin, den Eindruck von Julies relativem akademischen Leistungsniveau in den entsprechenden Notendurchschnitt zu übersetzen.

Der Intensitätsabgleich liefert Vorhersagen, die so extrem sind wie die Informationen, auf denen sie beruhen, und dies veranlasst Menschen dazu, auf zwei recht verschiedene Fragen dieselbe Antwort zu geben:

> Was ist Julies Perzentilwert auf der Skala der Lesefähigkeits-Frühreife?
> Was ist Julies Perzentilwert auf der Notenskala?

Mittlerweile sollten Sie problemlos erkannt haben, dass all diese Operationen Merkmale von System 1 sind. Ich habe sie hier als eine geordnete Folge von Schritten aufgelistet, aber natürlich funktioniert die Ausbreitung der Aktivierung im assoziativen Gedächtnis nicht in dieser

Weise. Sie sollten sich einen Prozess der Aktivierungsausbreitung vorstellen, der zunächst durch die Informationen und die Frage ausgelöst wurde, auf sich selbst zurückwirkt und sich schließlich für die kohärenteste mögliche Lösung entscheidet.

Amos und ich baten einmal die Probanden eines Experiments, Beschreibungen von acht College-Erstsemestern zu beurteilen, die angeblich von einem Studienberater stammten, der Erstsemester befragte. Jede Beschreibung bestand aus fünf Adjektiven, wie in dem folgenden Beispiel:

> intelligent, selbstbewusst, belesen, fleißig, wissbegierig

Wir baten einige Probanden, zwei Fragen zu beantworten:

> Wie stark beeindruckt Sie diese Beschreibung im Hinblick auf die Studienleistungen?
>
> Wie viele Beschreibungen von Erstsemestern (in Prozent) würden Sie Ihrer Einschätzung nach stärker beeindrucken?

Die Fragen verlangen von Ihnen, die Anhaltspunkte zu bewerten, indem Sie sie mit der Norm vergleichen, die Sie von Studenteneinschätzungen durch Studienberater haben. Schon die bloße Existenz einer solchen Norm ist bemerkenswert. Auch wenn Sie bestimmt nicht wissen, wie Sie diese erworben haben, haben Sie ein recht genaues Gespür dafür, wie viel Begeisterung in der Beschreibung mitschwingt: Der Studienberater ist der Ansicht, dass es sich um einen guten, aber keinen außergewöhnlich guten Studenten handelt. Er hätte stärkere Adjektive verwenden können als »intelligent« (»brillant«, »kreativ«), »belesen« (»gelehrt«, »hochgebildet«, »außerordentlich bewandert«) und »fleißig« (»leidenschaftlich«, »perfektionistisch«). Das Urteil: Sehr wahrscheinlich im Bereich der obersten 15 Prozent, aber eher nicht unter den obersten 3 Prozent. Bei

solchen Urteilen gibt es einen eindrucksvollen Konsens, zumindest innerhalb einer Kultur.

Den übrigen Teilnehmern unseres Experiments stellten wir andere Fragen:

> Was ist Ihre Schätzung für den Notendurchschnitt, den der Student erzielen wird?
>
> Wie hoch ist der Prozentsatz der Erstsemester, die einen besseren Notendurchschnitt erreichen?

Sie müssen genau hinsehen, um den subtilen Unterschied zwischen den beiden Fragenpaaren zu erkennen. Der Unterschied sollte offensichtlich sein, ist es aber nicht. Anders als die ersten Fragen, bei denen Sie nur die Informationen beurteilen mussten, ist das zweite Paar mit sehr viel mehr Ungewissheit behaftet. Die Frage bezieht sich auf die tatsächliche Leistung am Ende des ersten Studienjahres. Was geschah in dem Jahr, das vergangen ist, seitdem das Interview geführt wurde? Wie präzise können Sie die tatsächlichen Leistungen des Studenten im ersten College-Jahr auf der Basis von fünf Adjektiven vorhersagen? Würde der Berater selbst auf der Basis des Interviews den Notendurchschnitt absolut zutreffend vorhersagen?

Das Ziel dieser Studie war es, die Perzentil-Einschätzungen der Probanden, wenn sie in einem Fall die Informationen bewerteten und im anderen das tatsächliche Ergebnis vorhersagten, miteinander zu vergleichen. Die Ergebnisse lassen sich leicht zusammenfassen: Die Urteile waren identisch. Obgleich sich die beiden Fragenpaare voneinander unterschieden (in dem einen ging es um die Beschreibung, in dem anderen um die zukünftige akademische Leistung des Studenten), behandelten die Teilnehmer sie so, als wären sie gleich. Wie im Fall von Julie wird kein Unterschied gemacht zwischen der Vorhersage eines zukünftigen Ereignisses und der Beurteilung der gegebenen Informationen. Vorhersage und Bewertung decken sich. Dies ist vielleicht der beste Beleg für die Rolle der Ersetzung. Menschen werden um eine Vorhersage gebeten, aber sie ersetzen diese durch eine Bewertung der Information, ohne zu bemerken, dass die Frage, die sie beantworten, nicht die gleiche ist, die ihnen gestellt

wurde. Dieser Prozess führt zwangsläufig zu Vorhersagen, die mit einem systematischen Fehler behaftet sind; sie lassen völlig die Regression zum Mittelwert außer Betracht.

Während meines Wehrdienstes in der israelischen Armee war ich eine Zeit lang einer Einheit zugeteilt, die Bewerber für die Offizierslaufbahn auf Basis einer Reihe von Interviews und Praxistests auswählte. Das angegebene Kriterium für die erfolgreiche Vorhersage war die Abschlussnote eines Kandidaten auf der Offiziersschule. Die Gültigkeit der Einschätzungen war bekanntermaßen eher schlecht (ich werde in einem späteren Kapitel mehr dazu sagen). Die Einheit existierte auch noch Jahre später, als ich Professor war und mit Amos an der Studie über intuitive Urteile arbeitete. Ich hatte gute Kontakte zu den Leuten in dieser Einheit und bat sie um einen Gefallen. Zusätzlich zu dem üblichen Benotungssystem, das sie zur Einstufung der Anwärter benutzten, bat ich sie um ihre Einschätzung der Note, die jeder der zukünftigen Kadetten in der Offiziersschule erhalten würde. Sie sammelten einige Hundert dieser Vorhersagen. Die Offiziere, die die Vorhersagen machten, waren alle mit dem Benotungssystem auf Buchstabenbasis, das die Schule auf ihre Kadetten anwandte, und der ungefähren prozentualen Verteilung der Noten A, B und so weiter unter ihnen vertraut. Die Ergebnisse waren verblüffend: Die relative Häufigkeit der einzelnen Noten in den Vorhersagen entsprach fast genau deren Häufigkeiten unter den Endnoten der Schule.

Die Ergebnisse sind ein schlagendes Beispiel sowohl für Ersetzung wie für Intensitätsabstimmung. Die Offiziere, die die Vorhersagen machten, unterschieden nicht zwischen zwei Aufgaben:

– ihrer gewohnten Aufgabe, die darin bestand, die Leistung der Kandidaten während ihres Aufenthalts in der Einheit zu beurteilen
– der Aufgabe, die ich ihnen gestellt hatte, nämlich eine zukünftige Note vorherzusagen

Sie hatten einfach mithilfe der Intensitätsabstimmung ihre eigenen Einstufungen in die Notenskala der Offiziersschule übersetzt. Da sie sich – einmal mehr – nicht mit der (erheblichen) Unsicherheit ihrer Vorhersagen auseinandersetzen wollten, machten sie Vorhersagen, die die Regression zum Mittelwert völlig außer Acht ließen.

## Eine Korrektur für intuitive Vorhersagen

Kommen wir zurück zu Julie, unserer frühreifen Leserin. Die richtige Methode zur Vorhersage ihres Notendurchschnitts wurde im vorangehenden Kapitel eingeführt. Wie ich es dort für das Golfspielen an aufeinanderfolgenden Tagen und für Gewicht und Klavierspielen getan habe, schreibe ich hier eine schematische Formel für die Faktoren auf, die das Lesealter und die College-Noten determinieren:

Lesealter
= gemeinsame Faktoren + für das Lesealter spezifische Faktoren
= 100 Prozent

Notendurchschnitt
= gemeinsame Faktoren + für den Notendurchschnitt spezifische Faktoren
= 100 Prozent

Zu den gemeinsamen Faktoren zählen die genetisch determinierte Veranlagung, das Ausmaß, in dem die Familie akademische Interessen fördert, und alles andere, was dieselben Personen dazu veranlasst, als Kinder zu frühreifen Lesern zu werden und als junge Erwachsene einen guten Studienabschluss zu machen. Selbstverständlich gibt es viele Faktoren, die eines dieser Ergebnisse, nicht aber das andere beeinflussen. Julie könnte durch übermäßig ehrgeizige Eltern dazu gedrängt worden sein, früh mit dem Lesen zu beginnen, sie hatte vielleicht eine unglückliche Liebesgeschichte, die ihre College-Noten drückte, sie könnte in ihrer Jugend einen Skiunfall gehabt haben, der eine leichte Behinderung bei ihr hinterließ, und so weiter. Erinnern wir uns daran, dass die Korrelation zwischen zwei Messgrößen – in diesem Fall zwischen dem Lesealter und

dem Notendurchschnitt – gleich dem prozentualen Anteil der gemeinsamen Faktoren unter ihren Determinanten ist. Wie hoch schätzen Sie diesen Anteil? Meine optimistischste Schätzung liegt bei etwa 30 Prozent. Damit haben wir nun alles, was wir brauchen, um eine nicht verzerrte Vorhersage zu erzeugen. Folgen Sie den nachstehenden Anweisungen, um in vier einfachen Schritten dorthin zu gelangen:

1. Beginnen Sie mit einer Schätzung des mittleren Notendurchschnitts.
2. Legen Sie den individuellen Notendurchschnitt fest, der Ihrem Eindruck aus den vorliegenden Informationen entspricht.
3. Schätzen Sie die Korrelation zwischen Ihren Informationen und dem Notendurchschnitt.
4. Wenn die Korrelation 0,30 beträgt, bewegen Sie sich um 30 Prozent der Distanz vom allgemeinen Durchschnittswert zum passenden individuellen Notendurchschnitt.

Schritt 1 gibt Ihnen den allgemeinen Bezugswert, den Notendurchschnitt, den Sie vorhersagen würden, wenn Sie nichts weiter über Julie wüssten, außer dass sie im letzten Studienjahr ist und ihren Abschluss macht. Ohne weitere Informationen hätten Sie den Durchschnittswert vorhergesagt. (Das entspricht der Zuschreibung einer Basisraten-Wahrscheinlichkeit an BWL-Absolventen, wenn Sie keine weiteren Erkenntnisse über Tom W. haben.) Schritt 2 ist Ihre intuitive Vorhersage, die Ihrer Einschätzung der Informationen entspricht. Schritt 3 korrigiert den allgemeinen Bezugswert in Richtung Ihrer Intuition, aber die Entfernung, die Sie in Richtung der Intuition zurücklegen dürfen, hängt von Ihrer Schätzung der Korrelation ab. In Schritt 4 gelangen Sie schließlich zu einer Vorhersage, die zwar von Ihrer Intuition beeinflusst ist, aber sehr viel moderater als diese ausfällt.[1]

Dies ist eine allgemeingültige Vorhersagemethode. Sie können sie immer dann anwenden, wenn Sie eine quantitative Variable wie etwa den Notendurchschnitt vorhersagen müssen. Diese Methode geht von Ihrer Intuition aus, mäßigt sie aber durch eine Regression zum Mittelwert. Wenn Sie gute Gründe haben, um auf die Genauigkeit Ihrer intuitiven Vorhersage zu vertrauen – eine hohe Korrelation zwischen den Informationen und der Vorhersage –, müssen Sie nur geringfügig korrigieren.

Intuitive Vorhersagen müssen korrigiert werden, weil sie nicht regressiv und deshalb verzerrt sind. Angenommen, ich sage für jeden Golfer in einem Turnier vorher, dass sein Score an Tag zwei genau seinem Score an Tag eins

entsprechen wird. Diese Vorhersage berücksichtigt nicht die Regression zum Mittelwert: Die Golfer, die am ersten Tag gut abgeschnitten haben, werden, im Schnitt, am zweiten Tag eine schlechtere Leistung bringen, und diejenigen, die schlecht abgeschnitten haben, werden sich am meisten verbessern. Beim Vergleich mit den tatsächlichen Ergebnissen zeigt sich dann, dass nicht regressive Vorhersagen verzerrt sind. Sie sind im Schnitt allzu optimistisch für diejenigen, die am ersten Tag die besten Leistungen zeigten, und übermäßig pessimistisch für diejenigen, die einen schlechten Einstieg erwischt hatten. Die Vorhersagen sind genauso extrem wie die Informationen. Und wenn Sie anhand frühkindlicher Leistungen die College-Noten vorhersagen, ohne Ihre Vorhersagen durch eine Regression zum Mittelwert zu korrigieren, werden Sie von den Studienleistungen der Frühleser enttäuscht sein, während die Noten derjenigen, die relativ spät lesen lernten, Sie positiv überraschen werden. Die korrigierten intuitiven Vorhersagen beseitigen diese Verzerrungen, sodass (sowohl hohe als auch niedrige) Vorhersagen mit ungefähr gleich großer Wahrscheinlichkeit den wahren Wert über- beziehungsweise unterschätzen. Sie machen nach wie vor Fehler, auch wenn Ihre Vorhersagen nicht verzerrt sind, aber die Fehler sind kleiner und begünstigen weder hohe noch niedrige Ergebnisse.

## Eine Verteidigung extremer Vorhersagen?

Ich habe weiter vorn Tom W. eingeführt, um Vorhersagen diskreter (nicht kontinuierlicher) Ergebnisse wie Studienfach oder Prüfungserfolg zu veranschaulichen, die durch Zuweisung einer Wahrscheinlichkeit zu einem bestimmten Ereignis ausgedrückt werden (oder, in diesem Fall, durch die Anordnung der Ergebnisse nach der Reihenfolge ihrer Wahrscheinlichkeit). Ich beschrieb auch ein Verfahren, das den üblichen Verzerrungen diskreter Vorhersagen entgegenwirkt: Vernachlässigung der Basisraten und unzureichende Berücksichtigung der Qualität der Information.

Die Verzerrungen, die wir bei Vorhersagen, welche auf einer Skala ausgedrückt werden – wie etwa des Notendurchschnitts oder des Umsatzerlöses eines Unternehmens – finden, sind vergleichbar mit den Verzerrungen, die bei der Beurteilung der Wahrscheinlichkeiten von Ergebnissen beobachtet werden.

Die Korrekturverfahren sind ebenfalls ähnlich:

– Beide enthalten einen Vorhersage-Referenzwert *(baseline)*, an den man sich halten würde, wenn man nichts über den konkreten Fall wüsste. Im kategorialen Fall war dies die Basisrate. Im numerischen Fall ist es das durchschnittliche Ergebnis in der relevanten Kategorie.
– Beide enthalten eine intuitive Vorhersage, die die Zahl ausdrückt, die Ihnen spontan einfällt, egal ob es eine Wahrscheinlichkeit oder ein Notendurchschnitt ist.
– In beiden Fällen streben Sie nach einer Vorhersage, die zwischen dem Referenzwert und der intuitiven Antwort liegt.
– Wenn keine verwertbaren Informationen vorliegen, halten Sie sich an den Referenzwert.
– Im anderen Extremfall bleiben Sie bei Ihrer anfänglichen Vorhersage. Dies allerdings nur, wenn Sie nach einer kritischen Prüfung der Informationen, die Ihre anfängliche Vorhersage unterstützen, uneingeschränkt von dieser überzeugt bleiben.
– In den meisten Fällen finden Sie einen Grund dafür, zu bezweifeln, dass Ihr intuitives Urteil perfekt mit der Wahrheit korreliert, und Sie landen irgendwo zwischen den beiden Polen.

Dieses Verfahren ist eine Näherung der wahrscheinlichen Ergebnisse einer geeigneten statistischen Analyse. Wenn es gelingt, bringt es Sie unverzerrten Vorhersagen, sachgerechten Wahrscheinlichkeitseinschätzungen und angemessenen Prognosen numerischer Ergebnisse näher. Beide Verfahren sollen derselben Verzerrung entgegenwirken: Intuitive Vorhersagen basieren auf einer Selbstüberschätzung und sind zu extrem.

Die Korrektur unserer intuitiven Vorhersagen ist eine Aufgabe von System 2. Es bedarf erheblicher Anstrengungen, um die relevante Bezugskategorie zu finden, die Referenzprognose abzuschätzen und die Qualität der Information zu beurteilen. Diese Mühe ist nur dann gerechtfertigt, wenn viel auf dem Spiel steht und wenn es Ihnen besonders wichtig ist, keine Fehler zu machen. Außerdem sollten Sie wissen, dass es möglicherweise Ihr Leben komplizierter machen wird, wenn Sie Ihre Intuitionen korrigieren. Es ist ein Kennzeichen unverzerrter Vorhersagen, dass sie nur dann die Prognose seltener oder extremer Ereignisse erlauben, wenn die Informationen sehr gut sind. Wenn Sie erwarten, dass Ihre Vorhersagen nur von mittlerer Gültigkeit sind, werden Sie kein Ergebnis prognostizieren, das entweder selten oder weit weg vom Mittelwert ist. Wenn Ihre Vorhersagen nicht verzerrt sind, werden Sie nie die befriedigende Erfahrung machen, einen Extremfall zutreffend vorhergesehen zu haben. Sie werden nie »Das hab ich mir gedacht!« sagen können, wenn Ihr bester Jura-Student zum Richter am Obersten Gerichtshof ernannt wird oder wenn ein neu gegründetes Unternehmen, das Sie für sehr vielversprechend hielten, tatsächlich einen kometenhaften Aufstieg in seiner Branche durchmacht. Angesichts der Beschränktheit der verfügbaren Informationen werden Sie nie vorhersagen, dass ein hervorragender Highschool-Schüler ein Einser-Student an der Universität Princeton wird. Aus demselben Grund wird eine Wagniskapitalgesellschaft nie zu hören bekommen, die Erfolgsaussichten eines neu gegründeten Unternehmens in der Anfangsphase seien »sehr hoch«.

Die Kritik an dem Grundsatz, intuitive Vorhersagen zu mäßigen, muss ernst genommen werden, weil es nicht immer am wichtigsten ist, dafür zu sorgen, dass Verzerrungen beseitigt werden. Eine Präferenz für unverzerrte Vorhersagen ist dann gerechtfertigt, wenn sämtliche Vorhersagefehler gleich behandelt werden, ganz egal in welche Richtung sie gehen. Aber es gibt

Situationen, in denen ein bestimmter Fehler viel schlimmer ist als ein anderer. Wenn ein Risikokapitalgeber Ausschau hält nach »der nächsten großen Sache«, ist das Risiko, das nächste Google oder Facebook zu verpassen, viel wichtiger als das Risiko, eine bescheidene Summe in ein neues Unternehmen zu investieren, das schließlich am Markt scheitert. Risikokapitalgeber wollen die Extremfälle richtig vorhersagen, auch wenn sie dafür in Kauf nehmen müssen, die Erfolgsaussichten vieler anderer neu gegründeter Unternehmen zu überschätzen. Für eine vorsichtige Bank, die große Kredite vergibt, mag das Risiko, dass ein einzelner Kreditnehmer pleitegeht, schwerer wiegen als das Risiko, mehrere potenzielle Kunden abzulehnen, die ihre Verpflichtungen erfüllen würden. In diesen Fällen mögen extreme Formulierungen (»sehr gute Erfolgsaussichten«, »ernstes Ausfallrisiko«) wegen des Trostes, den sie vermitteln, eine gewisse Berechtigung haben, auch wenn die Informationen, auf denen diese Urteile basieren, nur von beschränkter Gültigkeit sind.

Für eine rationale Person sollten nicht verzerrte, maßvolle Vorhersagen kein Problem darstellen. Schließlich weiß die rationale Risikokapitalgeberin, dass selbst die vielversprechendsten Existenzgründungen nur eine mäßige Überlebenschance haben. Sie sieht ihre Aufgabe darin, aus den verfügbaren Wetten die vielversprechendsten auszuwählen, und sie hat nicht das Verlangen, sich über die Aussichten eines neu gegründeten Unternehmens, in das sie investieren will, Illusionen zu machen. In ähnlicher Weise werden rationale Personen, die den Umsatz einer Firma vorhersagen wollen, sich nicht auf eine bestimmte Zahl fokussieren – vielmehr betrachten sie den Unsicherheitsbereich um das wahrscheinlichste Ergebnis. Eine rationale Person investiert eine große Summe in ein Unternehmen, das mit hoher Wahrscheinlichkeit scheitert, wenn die materiellen Anreize im Erfolgsfall hinreichend groß sind, ohne sich über die Erfolgschancen Illusionen zu machen. Doch wir sind nicht alle rational, und einige von uns brauchen vielleicht die Sicherheit verzerrter Schätzwerte, um nicht in eine Lähmung zu verfallen. Wenn Sie sich jedoch selbst etwas vormachen wollen, indem Sie extreme Vorhersagen akzeptieren, sollten Sie sich immer Ihrer Nachgiebigkeit gegen sich selbst bewusst sein.

Der vielleicht größte Nutzen der von mir vorgeschlagenen Korrekturverfahren besteht darin, dass sie Sie dazu zwingen, darüber nachzudenken, was Sie wissen. Ich werde ein Beispiel verwenden, das in

der akademischen Welt allgemein bekannt ist, aber die Parallelen zu anderen lebensweltlichen Sphären sind unmittelbar ersichtlich. Ein Fachbereich will einen jungen Professor berufen und den Bewerber auswählen, der die höchste wissenschaftliche Produktivität verspricht. Der Berufungsausschuss hat die Auswahl auf zwei Kandidaten eingegrenzt:

> Kim hat vor Kurzem ihre Dissertation abgeschlossen. Ihre Empfehlungen sind spektakulär, sie hielt einen brillanten Vortrag und beeindruckte alle Anwesenden in ihren Vorstellungsgesprächen. Sie hat bislang allerdings nicht besonders viel publiziert.

> Jane hatte in den letzten drei Jahren eine Stelle als promovierte Nachwuchswissenschaftlerin inne. Sie war sehr produktiv, und sie hat exzellente Forschungsleistungen aufzuweisen, aber ihr Vortrag und ihre Vorstellungsgespräche waren nicht so bestechend wie die von Kim.

Intuitiv würde man sich für Kim entscheiden, weil sie einen stärkeren Eindruck hinterließ und wegen der Regel, dass nur die jeweils verfügbare Information zählt (WYSIATI). Aber es ist auch so, dass wir viel weniger Informationen über Kim als über Jane besitzen. Wir sind wieder zurück beim Gesetz der kleinen Zahlen. Tatsächlich verfügen wir über eine viel kleinere Stichprobe an Informationen über Kim als über Jane, und extreme Ergebnisse werden mit viel höherer Wahrscheinlichkeit in kleinen Stichproben beobachtet. Bei den Ergebnissen von kleinen Stichproben spielt der Zufall eine viel größere Rolle, und deshalb sollten wir bei unserer Vorhersage der zukünftigen Leistung Kims eine stärkere Regression zum Mittelwert vornehmen. Wenn wir die Tatsache berücksichtigen, dass Kim vermutlich stärker zum Mittelwert regrediert als Jane, werden wir vielleicht doch Jane auswählen, auch wenn sie uns weniger stark beeindruckt hat. Im Rahmen eines akademischen Berufungsverfahrens würde ich für Jane stimmen, aber ich müsste mühsam meinen intuitiven Eindruck überwinden,

dass Kim vielversprechender ist. Es ist natürlicher und irgendwie auch angenehmer, seinen Intuitionen zu folgen, als gegen sie zu handeln.

Sie können sich leicht ähnliche Probleme in anderen Zusammenhängen vorstellen, wie etwa ein Risikokapitalgeber, der sich entscheiden muss, in welche von zwei Neugründungen, die auf verschiedenen Märkten tätig sind, er investieren soll. Eines der neuen Unternehmen bietet ein Produkt an, für das sich recht genaue Nachfrageschätzungen erstellen lassen. Der andere Kandidat ist aufregender und intuitiv vielversprechender, aber seine Aussichten sind ungewisser. Ob die Erfolgsaussichten des zweiten Start-ups noch immer höher eingeschätzt werden, wenn die Unsicherheit berücksichtigt wird, ist eine Frage, die eine sorgfältige Erwägung verdient.

## Die Regression im Zwei-Systeme-Modell

Extreme Vorhersagen und die Bereitschaft, seltene Ereignisse aus schwachen Informationen vorherzusagen, sind beides Kennzeichen von System 1. Die assoziative Maschinerie setzt die Extremität der Vorhersagen automatisch mit der wahrgenommenen Extremität der Informationen gleich, auf denen diese basieren – so funktioniert die Ersetzung. Und System 1 erzeugt spontan Urteile, deren Gültigkeit wir überschätzen, weil der Grad des subjektiven Überzeugtseins, wie wir sahen, von der Kohärenz der besten Geschichte abhängt, die wir aus den verfügbaren Informationen konstruieren können. Seien Sie gewarnt: Ihre Intuitionen liefern Vorhersagen, die zu extrem sind, und Sie werden dazu neigen, ihnen allzu großen Glauben zu schenken.

Regression ist auch ein Problem für System 2. Schon die Idee der Regression zum Mittelwert ist befremdlich und schwer zu kommunizieren und zu verstehen. Galton musste sich sehr anstrengen, bis er sie verstand. Viele Statistiklehrer graut es vor der Stunde, in der dieses Thema behandelt wird, und ihre Schüler verstehen dieses so wichtige Konzept oftmals nur vage. Dies ist ein Fall, in dem System 2 ein spezielles Training braucht. Die Angleichung der Vorhersagen an die vorhandenen Informationen ist nicht nur etwas, was wir intuitiv tun; es scheint auch das zu sein, was vernünftig ist. Wir werden die Regression nicht durch Erfahrung verstehen lernen. Selbst wenn eine Regression erkannt wird, wie wir in der Geschichte der Fluglehrer sahen, wird ihr eine kausale Interpretation gegeben, die fast immer falsch ist.

### Zum Thema »Intuitive Vorhersagen«

> »Diesem Start-up gelang ein hervorragender Machbarkeitsnachweis, aber wir sollten nicht erwarten, dass es in der Zukunft genauso erfolgreich sein wird. Sie sind noch

weit vom Markt entfernt, und es gibt viel Spielraum für eine Regression.«

»Unsere intuitive Vorhersage ist sehr positiv, aber sie liegt vermutlich zu hoch. Wir sollten die Stärke unserer Daten berücksichtigen und die Vorhersage Richtung Mittelwert herabsetzen.«

»Die Investition in die Firma ist vielleicht eine gute Idee, auch wenn sie nach der besten Schätzung scheitern wird. Wir glauben nicht wirklich daran, dass sie das nächste Google wird.«

»Ich las eine Rezension dieser Marke, und sie war ausgezeichnet. Dennoch könnte es ein Glückstreffer gewesen sein. Wir sollten nur die Marken in Betracht ziehen, die viele Rezensionen bekommen haben, und diejenige auswählen, die am besten erscheint.«

# TEIL III

## Selbstüberschätzung

# *19. Die Illusion des Verstehens*

Der Wertpapierhändler, Philosoph und Statistiker Nassim Taleb könnte auch noch als Psychologe angesehen werden. In seinem Buch *Der Schwarze Schwan* führte Taleb den Begriff der »narrativen Verzerrung« *(narrative fallacy)* ein, um zu beschreiben, wie fehlerhafte Geschichten über die Vergangenheit unsere Weltanschauungen und Zukunftserwartungen prägen.[1] Narrative Verzerrungen gehen zwangsläufig aus unserem anhaltenden Bestreben hervor, die Welt zu verstehen. Die erklärenden Geschichten, die wir überzeugend finden, sind einfach; sie sind eher konkret als abstrakt und schreiben Begabung, Dummheit und Absichten eine größere Bedeutung zu als Glück. Sie konzentrieren sich auf ein paar markante Ereignisse, die geschehen sind, und nicht auf die zahllosen Ereignisse, die nicht geschehen sind. Jedes hervorstechende Ereignis aus jüngerer Zeit kann zum Kern einer kausalen Erzählung werden. Taleb behauptet, wir Menschen würden uns ständig selbst beschwindeln, indem wir fadenscheinige Berichte über die Vergangenheit konstruieren und sie für wahr halten.

Gute Geschichten bestehen aus einer einfachen und kohärenten Darstellung der Handlungen und Intentionen von Menschen. Wir neigen dazu, Verhaltensweisen als Manifestationen allgemeiner Neigungen und Persönlichkeitszüge zu interpretieren – Ursachen, die sich mühelos Wirkungen zuordnen lassen. Der früher diskutierte Halo-Effekt fördert die Kohärenz, weil er uns dazu veranlasst, sämtliche Eigenschaften einer Person aus unserer Sicht mit unserem Urteil über ein besonders bedeutsames Attribut zur Deckung zu bringen. [2] Wenn wir beispielsweise einen Baseballwerfer für attraktiv und sportlich halten, werden wir ihn wahrscheinlich auch als einen überdurchschnittlich guten Werfer einstufen.[3] Halo-Effekte können auch negativ sein: Wenn wir einen Spieler hässlich finden, werden wir seine sportlichen Fähigkeiten wahrscheinlich unterschätzen. Der Halo-Effekt trägt dazu bei, Erklärungen einfach und kohärent zu halten, indem er uns die Konsistenz von Beurteilungen überschätzen lässt: Gute Menschen tun nur Gutes, und schlechte Menschen

sind von Grund auf schlecht. Die Aussage »Hitler liebte Hunde und kleine Kinder« ist schockierend, ganz gleich wie oft man sie hört, weil jede Spur von Freundlichkeit bei jemandem, der so böse war, die vom Halo-Effekt erzeugten Erwartungen verletzt. Inkonsistenzen vermindern die Leichtigkeit unseres Denkens und die Klarheit unserer Gefühle.

Eine bezwingende Erzählung fördert die Illusion der Zwangsläufigkeit. Betrachten wir die Geschichte darüber, wie Google zum Giganten der Internetbranche wurde. Zwei kreative Studenten am Fachbereich Informatik der Universität Stanford entwickeln ein überlegenes Verfahren zur Informationssuche im Internet. Sie beantragen und erhalten Gelder zur Gründung eines Unternehmens und treffen eine Reihe von Entscheidungen, die sich als richtig erweisen. Innerhalb weniger Jahre wird das von ihnen gegründete Unternehmen zu einem der wertvollsten US-Unternehmen, und die beiden ehemaligen Studenten gehören zu den reichsten Menschen der Welt. Bei einer denkwürdigen Gelegenheit hatten sie Glück, was die Geschichte noch packender macht: Ein Jahr nach der Gründung von Google wollten sie ihre Firma für weniger als eine Million Dollar verkaufen, aber dem Käufer war der Preis zu hoch.[4] Die Erwähnung dieser einen glücklichen Begebenheit macht es leichter, die vielfältigen glücklichen Umstände, die das Ergebnis beeinflussten, zu unterschätzen.

Eine detaillierte Geschichte würde die Entscheidungen der Google-Gründer einzeln aufführen, aber für unsere Zwecke genügt die Feststellung, dass fast jede von ihnen getroffene Entscheidung zu einem positiven Ergebnis führte. Eine vollständigere Erklärung würde die Handlungen der Firmen beschreiben, die Google aus dem Feld schlug. Die glücklosen Konkurrenten würden als blind, träge und insgesamt überfordert mit der Bewältigung dieser Bedrohung, die sie schließlich aus dem Markt drängte, erscheinen.

Ich habe die Geschichte bewusst vage erzählt, aber Sie verstehen, worauf ich hinauswill: Das hier ist eine sehr gute Geschichte. Detaillierter ausgearbeitet, könnte die Geschichte Ihnen das Gefühl vermitteln, zu verstehen, wieso Google erfolgreich war; sie würde Ihnen auch das Gefühl geben, dass Sie eine nützliche allgemeine Lektion darüber gelernt haben, was Unternehmen zum Erfolg führt. Leider spricht vieles dafür, dass Ihr Gefühl, durch die Google-Geschichte etwas verstanden und aus ihr gelernt zu haben, weitgehend eine Illusion ist. Der entscheidende Test für die Güte einer Erklärung ist die Frage, ob sie das Ereignis im Vorhinein vorhersagbar

gemacht hätte. Keine Geschichte über Googles unwahrscheinlichen Erfolg wird diesen Test bestehen, weil keine Geschichte die zahllosen Ereignisse einbeziehen kann, die auch ein anderes Ergebnis hätten hervorbringen können. Das menschliche Gehirn beschäftigt sich nicht mit Nichtereignissen. Die Tatsache, dass viele der stattgefundenen wichtigen Ereignisse auf Entscheidungen beruhten, verleitet Sie noch stärker, die Rolle von Können und Geschick überzubewerten und den Anteil, den das Glück an dem Ergebnis hatte, zu unterschätzen. Weil jede maßgebliche Entscheidung positiv ausging, deutet die Geschichte auf ein beinahe makelloses Vorauswissen hin – aber Pech hätte jeden einzelnen der erfolgreichen Schritte zunichtemachen können. Der Halo-Effekt gibt dem Ganzen den letzten Schliff und verleiht den Helden der Geschichte eine Aura der Unbesiegbarkeit.

Wie das Beobachten eines geschickten Flößers, der beim Durchfahren der Stromschnellen einem gefährlichen Felsen nach dem anderen ausweicht, ist auch die Geschichte von Google wegen des ständigen Risikos des Scheiterns überaus spannend. Allerdings besteht ein aufschlussreicher Unterschied zwischen beiden Fällen. Der geschickte Flößer ist Hunderte Male durch die Stromschnellen gefahren. Er hat gelernt, das aufgewühlte Wasser vor sich zu lesen und Hindernisse vorwegzunehmen. Er hat gelernt, seine Körperhaltung ständig geringfügig zu korrigieren, um nicht umzufallen. Junge Männer haben weniger Gelegenheiten, zu lernen, wie man ein gigantisches Unternehmen gründet, und weniger Chancen, verborgene Felsen zu umschiffen – wie etwa die brillante Innovation einer Konkurrenzfirma. Natürlich spielten Können und Geschick eine große Rolle in der Erfolgsgeschichte von Google, aber Glück hatte in Wirklichkeit einen größeren Anteil daran, als in der Schilderung zum Ausdruck kommt. Und je größer der Anteil des Glücks, umso weniger lässt sich aus der Geschichte lernen.

Hier ist die mächtige WYSIATI-Regel am Werk. Wir können einfach nicht anders, als mit den beschränkten Informationen, die wir besitzen, so zu verfahren, als wären sie alles, was man über das Thema wissen kann. Aus den uns verfügbaren Informationen konstruieren wir die bestmögliche Geschichte, und wenn es eine gute Geschichte ist, glauben wir sie. Paradoxerweise ist es leichter, eine kohärente Geschichte zu entwerfen, wenn man wenig weiß, wenn man weniger Mosaiksteinchen zusammenfügen muss. Unsere beruhigende Überzeugung, dass die Welt

einen Sinn hat, ruht auf einem sicheren Fundament: unserer beinahe unbegrenzten Fähigkeit, die eigene Unwissenheit zu ignorieren.

Ich habe von zu vielen Leuten gehört, die »schon lange vor dem Ausbruch der Finanzkrise von 2008 wussten, dass diese unvermeidlich war«. Dieser Satz enthält ein höchst bedenkliches Wort, das wir bei Diskussionen über weitreichende Ereignisse aus unserem Vokabular streichen sollten. Dieses Wort ist »wussten«. Einige Leute glaubten lange im Voraus, dass es zu einer Krise kommen würde, aber sie wussten es nicht. Heute sagen sie, dass sie es wussten, weil sich die Krise tatsächlich ereignete. Dies ist eine falsche Verwendung eines wichtigen Begriffs. In der Alltagssprache verwenden wir das Wort »wissen« nur, wenn das, was gewusst wird, wahr ist, und zwar nachweislich wahr. Wir können nur dann etwas wissen, wenn es sowohl wahr als auch erkennbar ist. Aber diejenigen, die glaubten, dass es zu einer Krise kommen würde (und es gibt weniger davon, als sich heute daran erinnern, dies geglaubt zu haben), konnten es damals nicht schlüssig beweisen. Viele intelligente und gut informierte Menschen interessierten sich brennend für die Zukunft der Wirtschaft und glaubten nicht, dass eine Katastrophe unmittelbar bevorstand; aus dieser Tatsache folgere ich, dass die Krise nicht erkennbar war. Das Abwegige am Gebrauch des Wortes »wissen« in diesem Zusammenhang besteht nicht darin, dass einige Personen unverdientes Lob für eine Voraussicht einheimsen, die sie nicht hatten. Er besteht darin, dass die Verwendung dieses Wortes den Eindruck vermittelt, die Erkennbarkeit der Welt sei größer, als sie es tatsächlich ist. Sie trägt dazu bei, eine verderbliche Illusion aufrechtzuerhalten.

Der Kern dieser Illusion besteht darin, dass wir glauben, die Vergangenheit zu verstehen, woraus folgt, dass auch die Zukunft erkennbar sein sollte. In Wirklichkeit aber verstehen wir die Vergangenheit in geringerem Maße, als wir glauben. »Wissen« ist nicht das einzige Wort, das diese Illusion befördert. Im allgemeinen Sprachgebrauch sind auch die Wörter »Intuition« und »Vorahnung« für Gedanken der Vergangenheit reserviert, die sich als wahr erwiesen haben. Die Aussage »Ich hatte eine Vorahnung, dass die Ehe keinen Bestand haben würde, aber ich irrte mich« hört sich seltsam an, so wie es jeder Satz über eine Intuition tut, die sich als falsch erwiesen hat. Um klar durchdachte Aussagen über die Zukunft treffen zu können, müssen wir die Sprache bereinigen, die wir benutzen, um Überzeugungen der Vergangenheit zu benennen.

## Die sozialen Kosten der Rückschau

Das Gehirn, das sich Narrative über die Vergangenheit ausdenkt, ist ein sinnstiftendes Organ. Wenn ein unvorhergesehenes Ereignis eintritt, korrigieren wir unsere Sicht der Welt umgehend, um dieser Überraschung Rechnung zu tragen. Stellen Sie sich vor, es findet gleich ein Spiel zwischen zwei Footballmannschaften statt, die genau gleich viele Siege und Niederlagen aufzuweisen haben. Jetzt ist das Spiel vorbei, und eine Mannschaft hat die andere haushoch besiegt. In Ihrem korrigierten Weltmodell ist die siegreiche Mannschaft viel stärker als die Verliererin, und Ihre Sicht der Vergangenheit wie auch der Zukunft wurde durch diese neue Wahrnehmung verändert. Es ist vernünftig, aus Überraschungen zu lernen, aber es kann einige gefährliche Konsequenzen haben.

Eine allgemeine Beschränkung des menschlichen Geistes ist seine mangelhafte Fähigkeit, vergangene Wissenszustände oder Überzeugungen, die sich gewandelt haben, zu rekonstruieren. Sobald man sich eine neue Sicht der Welt (oder eines Teils von ihr) zu eigen macht, verliert man sofort einen Großteil seiner Fähigkeit, sich an das zu erinnern, was man glaubte, ehe man seine Einstellung änderte.

Viele Psychologen haben erforscht, was geschieht, wenn Menschen ihre Meinung ändern. Der Experimentator wählt dabei ein Thema aus, zu dem die Probanden sich noch keine abschließende Meinung gebildet haben – etwa die Todesstrafe –, und misst dann sorgfältig ihre Einstellungen. Als Nächstes sehen oder hören die Probanden eine Pro- oder Kontra-Botschaft. Nun misst der Experimentator erneut die Einstellungen der Versuchspersonen; es zeigt sich, dass sie sich in der Regel der überzeugenden Botschaft, die ihnen dargeboten wurde, angenähert haben. Schließlich sollen die Probanden ihre frühere Meinung erläutern. Diese Aufgabe erweist sich als erstaunlich schwierig. Statt ihre früheren Überzeugungen zu rekonstruieren, erinnern sich die Versuchspersonen an ihre gegenwärtigen – ein Beispiel für Ersetzung –, und viele können nicht glauben, dass sie jemals eine andere Meinung hatten.[5]

Unsere Unfähigkeit, frühere Überzeugungen zu rekonstruieren, veranlasst uns zwangsläufig dazu, das Ausmaß zu unterschätzen, in dem wir durch vergangene Ereignisse überrascht wurden. Baruch Fischhoff hat als

Student in Jerusalem als Erster den »Ich wusste es die ganze Zeit«-Effekt oder »Rückschaufehler« nachgewiesen. Zusammen mit Ruth Beyth (einer weiteren Studentin von uns) führte Fischhoff vor dem Besuch von Präsident Richard Nixon in China und Russland im Jahr 1972 eine Umfrage durch. Die Befragten sollten 15 möglichen Ergebnissen von Nixons diplomatischen Initiativen Wahrscheinlichkeiten zuschreiben. Wäre Mao Tsetung bereit, sich mit Nixon zu treffen? Würden die Vereinigten Staaten China diplomatisch anerkennen? Könnten sich die Vereinigten Staaten und die Sowjetunion nach jahrzehntelanger Feindschaft auf ein substanzielles Abkommen verständigen?[6]

Nach Nixons Rückkehr forderten Fischhoff und Beyth dieselben Personen auf, sich an die Wahrscheinlichkeit zu erinnern, die sie ursprünglich jedem der 15 möglichen Ergebnisse zugeschrieben hatten. Die Befunde waren eindeutig. Wenn ein Ereignis tatsächlich eingetreten war, überschätzten die Befragten die Wahrscheinlichkeit, die sie dem Ereignis früher zugeschrieben hatten. Wenn das mögliche Ereignis nicht eingetreten war, erinnerten sie sich fälschlich, dass sie es schon immer für unwahrscheinlich hielten. Weitere Experimente zeigten, dass Menschen die Genauigkeit nicht nur ihrer eigenen früheren Vorhersagen, sondern auch die von anderen überbewerteten. Ähnliche Ergebnisse wurden für andere Ereignisse gefunden, die große öffentliche Aufmerksamkeit erregten, wie etwa der Mordprozess gegen O. J. Simpson und das Amtsenthebungsverfahren gegen Präsident Bill Clinton. Die Neigung, die Geschichte der eigenen Überzeugungen im Lichte der tatsächlichen Ereignisse umzuschreiben, erzeugt eine robuste kognitive Illusion.

Der Rückschaufehler übt einen schädlichen Einfluss auf die Beurteilungen von Entscheidungsträgern aus. Er veranlasst Beobachter dazu, die Güte einer Entscheidung nicht danach zu beurteilen, ob der Prozess der Entscheidungsfindung fehlerfrei war, sondern danach, ob das Ergebnis positiv oder negativ war.[7] Betrachten wir einen risikoarmen chirurgischen Eingriff, bei dem sich ein unvorhergesehener Zwischenfall ereignete, der den Tod des Patienten verursachte. Die Geschworenen werden, in der Rückschau, zu der Ansicht neigen, dass die Operation von Anfang an riskant war und dass der Arzt, der sie anordnete, es hätte besser wissen müssen. Dieser Ergebnisfehler *(outcome bias)* macht es fast unmöglich, eine Entscheidung sachgerecht zu beurteilen – im Hinblick auf

die Überzeugungen, die vernünftig waren, als die Entscheidung getroffen wurde.

Die Rückschau ist besonders unfreundlich zu Entscheidern, die als Vertreter der Interessen anderer handeln – Ärzte, Finanzberater, Trainer, Topmanager, Sozialarbeiter, Diplomaten, Politiker. Wir neigen dazu, Entscheidungsträger für gute Entscheidungen, die einen negativen Ausgang nehmen, zu tadeln, und für erfolgreiche Maßnahmen, die erst im Nachhinein naheliegend erscheinen, nicht genug zu loben. Es gibt einen eindeutigen Ergebnisfehler. Wenn die Ergebnisse schlecht sind, machen die Klienten oftmals ihre Vertreter dafür verantwortlich, dass sie die Zeichen an der Wand nicht gesehen haben – wobei sie vergessen, dass das Menetekel mit unsichtbarer Tinte geschrieben war, die erst im Nachhinein lesbar wurde. Handlungen, die im Vorfeld besonnen wirken, können in der Rückschau in unverantwortlicher Weise nachlässig erscheinen. Basierend auf einem realen Rechtsfall, wurden Studenten in Kalifornien gefragt, ob die Stadt Duluth in Minnesota die erheblichen Kosten für die Anstellung eines Vollzeit-Brückenaufsehers hätte schultern sollen, um sich gegen das Risiko abzusichern, dass sich Geröll verfängt und den freien Fluss des Wassers blockiert. Einer Gruppe wurden nur die zum Zeitpunkt der Entscheidung für die Stadt verfügbaren Informationen mitgeteilt; 24 Prozent dieser Probanden waren der Meinung, Duluth hätte die Ausgaben für einen solcher Aufseher auf sich nehmen sollen. Die zweite Gruppe wurde darüber informiert, dass Geröll den Fluss blockiert und schwere Flutschäden angerichtet hatte; 56 Prozent dieser Probanden sagten, die Stadt hätte den Aufseher einstellen sollen, obwohl sie ausdrücklich angewiesen worden waren, ihr Urteil nicht durch das nachträgliche Wissen verzerren zu lassen.[8]

Je schlimmer die Folgen, umso größer der Rückschaufehler. Im Fall einer Katastrophe wie etwa der Terroranschläge vom 11. September 2001 sind wir in besonders hohem Maße bereit, zu glauben, dass die Amtsträger, die sie nicht vorhersahen, nachlässig oder blind waren. Am 10. Juli 2001 erhielt die Central Intelligence Agency (CIA) Informationen, wonach El Kaida einen größeren Angriff gegen die Vereinigten Staaten plane. Der damalige CIA-Direktor George Tenet unterrichtete nicht Präsident George W. Bush, sondern die Nationale Sicherheitsberaterin Condoleezza Rice über diese Erkenntnisse. Als das später herauskam, schrieb Ben Bradlee, der legendäre Chefredakteur der *Washington Post*: »Es erscheint mir selbstverständlich,

dass man eine solche Nachricht, die Geschichte schreiben wird, direkt dem Präsidenten mitteilt.« Aber am 10. Juli wusste niemand – hätte niemand wissen können –, dass diese Neuigkeit Geschichte schreiben würde.[9]

Weil sich die Befolgung der üblichen Vorgehensweise nur schwer im Nachhinein kritisieren lässt, neigen Entscheidungsträger, die damit rechnen, dass ihre Entscheidungen im Nachhinein genau überprüft werden, zu bürokratischen Lösungen – und zu einer extremen Risikoscheu.[10] Als die Zahl der Kunstfehlerprozesse immer weiter zunahm, änderten Ärzte in vielfältiger Weise ihr Verhalten gegenüber Patienten: Sie gaben mehr Tests in Auftrag, sie überwiesen mehr Patienten an Spezialisten, sie wandten konventionelle Behandlungen an, auch wenn diese vermutlich nicht helfen würden. Diese Handlungen schützten die Ärzte stärker, als sie den Patienten nützten, und sie führten zu potenziellen Interessenkonflikten. Verstärkte Rechenschaftspflicht ist ein zweischneidiges Schwert.

Auch wenn Rückschau- und Ergebnisfehler im Allgemeinen die Risikoscheu fördern, bringen sie unverantwortlichen Hasardeuren auch unverdiente Belohnungen, wie etwa einem General oder einem Unternehmer, die ein aberwitziges Risiko eingingen, das sich auszahlte. Führungspersonen, die Glück haben, werden nicht dafür bestraft, dass sie zu hohe Risiken eingegangen sind. Stattdessen glaubt man, sie hätten das Gespür und den Weitblick besessen, um den Erfolg vorherzusehen, und die vernünftigen Leute, die an ihnen zweifelten, werden im Rückblick als mittelmäßig, scheu und schwach beurteilt. Ein paar glücklich ausgegangene Wagnisse können einen rücksichtslosen Führer mit einem Halo der Voraussicht und Kühnheit umgeben.

## Erfolgsrezepte

Die Sinngebungsmaschinerie von System 1 lässt uns die Welt geordneter, einfacher, vorhersagbarer und kohärenter sehen, als sie es tatsächlich ist. Die Illusion, man habe die Vergangenheit verstanden, nährt die weitere Illusion, man könne die Zukunft vorhersagen und kontrollieren. Diese Illusionen sind beruhigend. Sie verringern die Angst, die wir zu spüren bekämen, wenn wir uns die Ungewissheiten des Daseins uneingeschränkt bewusst machen würden. Wir alle haben ein Bedürfnis nach der beruhigenden Botschaft, dass Handlungen die gewünschten Folgen haben und dass Klugheit und Mut von Erfolg gekrönt sind. Viele Wirtschaftsbücher sind darauf zugeschnitten, dieses Bedürfnis zu befriedigen.

Beeinflussen Führungspersönlichkeiten und Führungspraktiken den Erfolg von Unternehmen am Markt? Natürlich tun sie das, und die Effekte wurden durch systematische Forschungen bestätigt, die die Merkmale von Vorstandschefs und ihrer Entscheidungen objektiv bewerteten und sie zu den späteren Erfolgen oder Misserfolgen des Unternehmens in Beziehung setzten. In einer Studie wurden die Topmanager durch die Strategie der Unternehmen, die sie vor ihrer gegenwärtigen Position geführt hatten, sowie durch die Führungsregeln und -verfahren, die nach ihrer Ernennung eingeführt wurden, charakterisiert. [11] Die Chefs beeinflussen den Unternehmenserfolg, aber der Effekt ist viel kleiner, als es die Lektüre der Wirtschaftspresse nahelegt.

Forscher messen die Stärke von Beziehungen mit einem Korrelationskoeffizienten, der zwischen null und eins schwankt. Der Koeffizient wurde weiter vorn (im Verhältnis zur Regression zum Mittelwert) durch das Ausmaß definiert, in dem zwei Messgrößen durch gemeinsame Faktoren determiniert werden. Eine sehr großzügige Schätzung der Korrelation zwischen dem Erfolg eines Unternehmens und der Qualität ihres Chefs könnte 0,30 betragen, was eine Überlappung von 30 Prozent anzeigt. Um zu verstehen, was diese Zahl bedeutet, betrachten Sie die folgende Frage:

> Angenommen, Sie untersuchen viele Paare von Firmen. Die beiden Firmen in jedem Paar sind im Allgemeinen ähnlich, aber der Chef der einen ist tüchtiger als der der anderen. Wie oft ist die Firma mit dem besseren Chef die erfolgreichere der beiden?[12]

In einer wohlgeordneten und vorhersagbaren Welt wäre die Korrelation perfekt (1,0), und der leistungsstärkere Chef würde in 100 Prozent der Paare das erfolgreichere Unternehmen führen. Wenn der relative Erfolg ähnlicher Firmen ausschließlich von Faktoren bestimmt würde, die der Chef nicht kontrolliert (man könnte sie, wenn man wollte, Glück nennen), würde das erfolgreichere Unternehmen in 50 Prozent der Fälle von dem schwächeren Chef geführt werden. Eine Korrelation von 0,30 bedeutet, dass der leistungsstärkere Chef bei etwa 60 Prozent der Paare die erfolgreichere Firma leitet – eine Verbesserung von nur 10 Prozent gegenüber einer Zufallsschätzung, was die geradezu heldenhafte Verehrung von Vorstandschefs, die wir so oft erleben, kaum rechtfertigt.

Wenn Sie einen höheren Wert erwartet hätten – und die meisten von uns tun das –, dann sollten Sie darin einen Hinweis sehen, dass Sie dazu neigen, die Vorhersagbarkeit der Sie umgebenden Welt zu überschätzen. Damit wir uns nicht missverstehen: Der Verbesserung der Erfolgsaussichten von 1 : 1 auf 3 : 2 ist ein sehr beachtlicher Vorteil, sowohl auf der Pferderennbahn als auch im Geschäftsleben. Aus der Sicht der meisten Wirtschaftspublizisten jedoch wäre ein Vorstandschef, der den Unternehmenserfolg nur in einem so geringen Ausmaß beeinflussen kann, keine sonderlich imponierende Führungskraft, selbst wenn sein Unternehmen prosperierte. Man kann sich nur schwer vorstellen, dass sich Menschen in Flughafenbuchhandlungen anstellen würden, um ein Buch zu kaufen, das euphorisch die Methoden von Topmanagern beschreibt, deren Leistungen im Schnitt nur geringfügig über der Zufallsrate liegen. Die Verbraucher lechzen nach einer eindeutigen Botschaft über die bestimmenden Faktoren von Erfolg und Misserfolg im Wirtschaftsleben, und sie brauchen Geschichten, die ihnen Sinnzusammenhänge vermitteln, und seien diese auch trügerisch.

In seinem scharfsinnigen Buch *Der Halo-Effekt* zeigt der in der Schweiz lehrende Wirtschaftsprofessor Philip Rosenzweig, wie das Bedürfnis nach trügerischer Gewissheit in zwei populären Genres der Managementliteratur befriedigt wird: Geschichten über den Aufstieg (der Regelfall) und den Fall (gelegentlich) bestimmter Personen und Unternehmen sowie Analysen der Unterschiede zwischen erfolgreichen und weniger erfolgreichen Firmen.[13] Sein Fazit lautet, dass Geschichten über Erfolg und Misserfolg durchgängig die Auswirkungen des Führungsstils und der Managementpraktiken auf den Geschäftserfolg von Unternehmen überschätzen, und daher ist ihre Botschaft nur selten nützlich.

Um zu verstehen, was hier abläuft, stellen Sie sich vor, dass Wirtschaftsexperten, etwa andere Vorstandschefs, gebeten werden, den Ruf des Chefs eines Unternehmens einzuschätzen. Sie wissen ganz genau, ob das Unternehmen in jüngster Zeit florierte oder schwächelte. Wie wir vorher im Fall von Google sahen, erzeugt dieses Wissen einen Halo-Effekt. Der Chef eines erfolgreichen Unternehmens wird wahrscheinlich als flexibel, methodisch und entscheidungsfreudig gelobt. Stellen Sie sich vor, es ist ein Jahr vergangen, und das Unternehmen ist auf dem absteigenden Ast. Derselbe Manager wird jetzt als orientierungslos, starr und autoritär beschrieben. Beide Beschreibungen hörten sich zu ihrem jeweiligen Zeitpunkt vernünftig an: Es wirkt geradezu absurd, eine erfolgreiche Führungskraft als starr und orientierungslos zu charakterisieren oder eine ums Überleben ringende Führungskraft als flexibel und methodisch zu bezeichnen.

Tatsächlich ist der Halo-Effekt so stark, dass Sie vermutlich die Auffassung ablehnen, dieselbe Person und dieselben Verhaltensweisen würden methodisch erscheinen, wenn alles gut läuft, und starr, wenn die Dinge schlecht laufen. Aufgrund des Halo-Effekts drehen wir die kausale Beziehung um: Wir neigen zu der Auffassung, dass Firmen scheitern, weil ihr Chef zu unbeweglich ist, während es in Wahrheit so ist, dass der Chef unbeweglich erscheint, weil die Geschäfte der Firma schlecht laufen. So entstehen Erkenntnisillusionen.

Der Halo-Effekt und der Ergebnisfehler erklären beide gemeinsam den außerordentlichen Erfolg von Büchern, die versuchen, aus der systematischen Untersuchung erfolgreicher Unternehmen konkrete Handlungsanweisungen abzuleiten. Eines der bekanntesten Beispiele dieses Genres ist das Buch *Immer erfolgreich* von Jim Collins und Jerry I. Porras.

Es enthält eine gründliche Analyse von 18 Paaren konkurrierender Unternehmen, bei denen eines erfolgreicher war als das andere. Die Daten für diese Vergleiche bestanden aus Beurteilungen verschiedener Aspekte der Unternehmenskultur und -strategie sowie der Managementpraktiken. »Wir glauben, dass jeder Vorstandschef, Manager und Unternehmer in der Welt dieses Buch lesen sollte«, verkünden die Autoren. »Sie können ein visionäres Unternehmen aufbauen.«[14]

Die Kernbotschaft von *Immer erfolgreich* und ähnlichen Büchern lautet, dass man gute Führungspraktiken identifizieren kann und dass gute Praktiken durch gute Ergebnisse belohnt werden. Beide Behauptungen sind überzogen. Der Vergleich von Firmen, die mehr oder weniger erfolgreich sind, ist in einem erheblichen Ausmaß ein Vergleich zwischen Firmen, die mehr oder minder viel Glück hatten. Wenn man weiß, wie wichtig der Faktor Glück ist, sollte man besonders argwöhnisch sein, wenn aus dem Vergleich von erfolgreichen und weniger erfolgreichen Firmen hochkonsistente Muster hervorgehen. Wenn der Zufall seine Hand im Spiel hat, können regelmäßige Muster nur Illusionen sein.

Weil Glück eine so große Rolle spielt, lassen sich aus empirischen Erfolgsdaten keine zuverlässigen Rückschlüsse auf die Qualität des Führungsteams und der Managementpraktiken ziehen. Und selbst wenn wir im Voraus absolut zuverlässig wüssten, dass ein Vorstandschef eine brillante Vision und außerordentliche Kompetenz besitzt, könnten wir noch immer nicht mit einer weit über das Zufallsergebnis eines Münzwurfs hinausgehenden Genauigkeit vorhersagen, wie erfolgreich das Unternehmen sein wird.[15] Im Schnitt schwand der Abstand in Ertragskraft und Aktienrendite zwischen den herausragenden und den weniger erfolgreichen Firmen, die in *Immer erfolgreich* untersucht wurden, in der Zeit nach Abschluss der Studie praktisch auf null. Die durchschnittliche Ertragskraft der in dem berühmten Buch *Auf der Suche nach Spitzenleistungen* aufgelisteten Unternehmen brach ebenfalls innerhalb kurzer Zeit stark ein.[16] Eine Studie über die vom Wirtschaftsmagazin *Fortune* auserkorenen »meistbewunderten Firmen« kam zu dem Ergebnis, dass über einen Zeitraum von zwanzig Jahren die Firmen mit den schlechtesten Bewertungen im weiteren Verlauf viel höhere Aktienrenditen erzielten als die meistbewunderten Unternehmen.[17]

Sie neigen wahrscheinlich zu kausalen Erklärungen dieser Beobachtungen: Vielleicht wurden die erfolgreichen Firmen allzu selbstzufrieden, und die weniger erfolgreichen strengten sich stärker an. Aber das ist die falsche Sicht. Der durchschnittliche Abstand muss zurückgehen, weil sich die anfängliche Kluft größtenteils dem Zufall verdankt, der sowohl zu dem Erfolg der Spitzenfirmen wie zu der Leistungsschwäche der anderen beitrug. Wir sind dieser statistischen Tatsache bereits begegnet: der Regression zum Mittelwert.[18]

Geschichten über den Aufstieg und Fall von Unternehmen stoßen bei Lesern auf Resonanz, weil sie das anbieten, was der menschliche Intellekt braucht: eine einfache Botschaft von Sieg und Niederlage, die eindeutige Ursachen identifiziert und die bestimmende Macht des Zufalls sowie die Unvermeidlichkeit einer statistischen Regression ausblendet. Diese Geschichten lösen eine Illusion des Verstehens aus und erhalten sie aufrecht, und sie vermitteln dem Leser, der sie nur allzu bereitwillig glaubt, Lektionen, die ihm keinen dauerhaften Nutzen bringen.

## Zum Thema »Rückschau«

»Der Fehler scheint offensichtlich zu sein, aber das zeigt sich erst im Rückblick. Im Voraus hätte man ihn nicht erkennen können.«

»Er glaubt dieser allzu glatten Erfolgsgeschichte zu sehr. Er ist einem narrativen Fehlschluss aufgesessen.«

»Sie hat keinen Beweis für die Behauptung, die Firma werde schlecht gemanagt. Sie weiß nur, dass der Aktienkurs eingebrochen ist. Dies ist ein Ergebnisfehler, teils Rückschau-Verzerrung, teils Halo-Effekt.«

»Wir sollten nicht auf den Ergebnisfehler hereinfallen. Dies war eine dumme Entscheidung, auch wenn sie gut ausging.«

# *20. Die Illusion der Gültigkeit*

System 1 ist so gestaltet, dass es aus dürftigen Informationen automatisch weitreichende Schlussfolgerungen zieht – es ist nicht imstande, zu ermessen, wie groß die Sprünge sind, die es beim Folgern macht. Wegen der WYSIATI-Regel zählen nur die verfügbaren Informationen. Weil der Grad der Überzeugung von der Kohärenz abhängt, spiegelt sich in dem subjektiven Überzeugtsein von unseren Meinungen die Kohärenz der Geschichte wider, die System 1 und System 2 entworfen haben. Die Menge an Informationen und deren Qualität spielen keine große Rolle, da sich auch aus dürftigen Informationen eine sehr gute Geschichte erarbeiten lässt. Für einige unserer wichtigsten Überzeugungen haben wir keinerlei Belege, außer dass Menschen, die wir mögen und denen wir vertrauen, diese Überzeugungen teilen. In Anbetracht dessen, wie wenig wir wissen, ist es absurd, wie fest wir an unsere Überzeugungen glauben – aber es ist auch unerlässlich.

# Die Illusion der Gültigkeit

Vor etlichen Jahrzehnten verbrachte ich sehr viel Zeit unter einer stechenden Sonne und beobachtete Gruppen verschwitzter Soldaten bei der Lösung eines Problems. Ich leistete damals meinen Wehrdienst in der israelischen Armee ab. Ich hatte gerade meinen ersten akademischen Grad in Psychologie erworben und wurde nach einem Jahr als Infanterie-Offizier zum psychologischen Dienst der Armee versetzt, wo eine meiner gelegentlichen Pflichten darin bestand, bei der Beurteilung von Offiziersanwärtern zu helfen. Wir benutzten dabei Methoden, die von der britischen Armee im Zweiten Weltkrieg entwickelt worden waren.

Ein Test, die sogenannte »Gruppenaufgabe ohne Anführer«, wurde auf einem Hindernisfeld durchgeführt. Acht Kandidaten, die sich untereinander nicht kannten und die anstelle von Rangabzeichen zur Identifizierung nur Nummernschildchen trugen, sollten einen langen Baumstamm vom Boden aufheben und zu einer etwa 1,80 Meter hohen Mauer tragen. Die gesamte Gruppe sollte über die Mauer steigen, ohne dass der Stamm den Boden oder die Mauer berührt und ohne dass eines der Gruppenmitglieder die Mauer berührt. Falls dies doch geschehen sollte, mussten die Kandidaten es melden und von vorn beginnen.

Es gab mehr als einen Lösungsweg für das Problem. Eine gängige Lösung bestand darin, dass mehrere Männer des Teams über den Pfahl auf die andere Seite krochen, während der Stamm von anderen Gruppenmitgliedern schräg gehalten wurde wie eine riesige Angel. Oder einige Soldaten kletterten auf die Schulter eines Kameraden und sprangen auf die andere Seite. Der letzte Mann musste dann auf den Pfahl hinaufspringen, der von dem Rest der Gruppe schräg gehalten wurde, längs über ihn klettern, während die anderen ihn und den Pfahl in der Luft hielten, und sicher auf die andere Seite springen. Hier kam es häufig zu Fehlern, sodass sie von vorn beginnen mussten.

Als ein Kollege und ich die Übung beobachteten, notierten wir, wer die Regie übernahm, wer zu führen versuchte, aber abgewiesen wurde, und wie kooperativ sich jeder Soldat an der gemeinsamen Anstrengung der Gruppe beteiligte. Wir sahen, wer störrisch zu sein schien, wer unterwürfig, arrogant, geduldig, jähzornig, beharrlich oder ein Drückeberger. Wir sahen

manchmal jemanden, dessen Vorschlag von der Gruppe abgelehnt worden war und der sich danach aus gekränkter Eitelkeit nicht mehr sonderlich ins Zeug legte. Und wir sahen, wie auf kritische Reaktionen reagiert wurde: Wer einen Kameraden beschimpfte, dessen Fehler das Scheitern der ganzen Gruppe nach sich zog, wer sich als Anführer hervortat, wenn das erschöpfte Team wieder von vorn anfangen musste. Unter dem Stress des Ereignisses, so schien uns, kam das wahre Naturell jedes Mannes zum Vorschein. Unser Eindruck vom Charakter jedes Anwärters war so klar und evident wie die Farbe des Himmels.

Nachdem wir mehrere Versuche der Kandidaten beobachtet hatten, mussten wir unsere Eindrücke von den Führungsfähigkeiten der Soldaten zusammenfassen und mit einem Punktwert bestimmen, wer zur Offiziersausbildung zugelassen werden sollte. Wir diskutierten jeden Fall gründlich und ließen dabei unsere Eindrücke Revue passieren. Die Aufgabe war nicht schwierig, weil wir glaubten, die Führungsfähigkeiten jedes Soldaten gut einschätzen zu können. Einige der Männer hatten sich scheinbar als starke Führungspersönlichkeiten hervorgetan, andere hatten den Eindruck von Feiglingen oder arroganten Dummköpfen gemacht, wieder andere hatten eine mittelmäßige Leistung gezeigt, schienen aber nicht hoffnungslos zu sein. Ziemlich viele machten einen so schwachen Eindruck, dass wir sie als Offiziersanwärter aussiebten. Als sich unsere vielfältigen Beobachtungen an jedem Kandidaten zu einer kohärenten Geschichte verdichteten, waren wir uneingeschränkt von unseren Bewertungen überzeugt und hatten das Gefühl, dass unsere Beobachtungen eine zuverlässige Prognose über das zukünftige Verhalten der Kandidaten erlaubten. Der Soldat, der die Regie übernahm, als die Gruppe in Schwierigkeiten war, und der das Team über die Mauer lotste, war in diesem Moment ein Anführer. Da drängte sich die Einschätzung auf, dass er sich in der Ausbildung oder im Kampf genauso bewähren würde wie bei der Gruppenaufgabe. Jede andere Vorhersage schien nicht mit unseren Beobachtungen in Einklang zu stehen.

Weil unsere Eindrücke von der Leistungsfähigkeit jedes Soldaten im Allgemeinen kohärent und eindeutig waren, fielen unsere formalen Prognosen genauso bestimmt aus. Ein gewisser Punktwert schien sich von selbst aufzudrängen, und wir hatten nur selten Zweifel oder widersprüchliche Eindrücke. Wir erklärten mit der größten Selbstverständlichkeit: »Dieser wird es nie schaffen«, »Dieser Kerl ist

mittelmäßig, sollte aber durchkommen« oder »Der wird ein Star«. Wir sahen keinerlei Anlass, unsere Prognosen infrage zu stellen, sie abzumildern oder einem eindeutigen Urteil auszuweichen. Bei kritischen Nachfragen waren wir jedoch bereit, zuzugeben: »Natürlich könnte alles passieren.« Wir waren bereit, dies einzuräumen, denn trotz unserer eindeutigen Eindrücke von den einzelnen Kandidaten wussten wir mit Sicherheit, dass unsere Prognosen weitgehend nutzlos waren.

Es gab überwältigende Belege dafür, dass wir den erfolgreichen Abschluss der Offiziersausbildung nicht präzise vorhersagen konnten. Alle paar Monate hatten wir eine Feedback-Sitzung, in der wir erfuhren, welche Leistungen die Kadetten auf der Offiziersschule zeigten, und unsere Beurteilungen mit den Einschätzungen der Kommandeure, die sie schon einige Zeit beobachteten, vergleichen konnten. Es war immer das Gleiche: Unsere Fähigkeit, die Leistung auf der Akademie vorherzusagen, war vernachlässigbar gering. Unsere Vorhersagen waren besser als blindes Raten – aber nicht sehr viel.

Nachdem wir die entmutigenden Neuigkeiten erfahren hatten, waren wir eine Zeit lang niedergeschlagen. Aber wir waren in der Armee. Ob nützlich oder nicht, es gab bestimmte Abläufe, die einzuhalten, und Befehle, die zu befolgen waren. Am nächsten Tag traf ein neuer Schwung von Kandidaten ein. Wir führten sie zu dem Hindernisfeld, wir konfrontierten sie mit der Mauer, sie hoben den Stamm auf, und innerhalb einiger Minuten glaubten wir so klar und deutlich wie zuvor ihre wahre Natur zu erkennen. Die betrübliche Wahrheit über die Qualität unserer Vorhersagen hatte keinerlei Auswirkung auf unsere Beurteilung der Kandidaten und nur sehr geringe Auswirkungen darauf, wie überzeugt wir von unseren Urteilen und Vorhersagen über Einzelpersonen waren.

Was passierte, war bemerkenswert. Die umfassenden Belege für unser vorhergehendes Versagen hätten unser Vertrauen in unsere Urteile über die Kandidaten erschüttern müssen, aber das geschah nicht. Sie hätten uns dazu veranlassen sollen, unsere Vorhersagen zu mäßigen, aber das taten sie nicht. Wir wussten, dass unsere Vorhersagen grundsätzlich kaum besser als ein blindes Raten waren, aber wir fühlten und verhielten uns weiterhin so, als wäre jede unserer konkreten Vorhersagen gültig. Ich musste an die Müller-Lyer-Illusion denken, bei der die Geraden gleich lang sind, aber trotzdem unterschiedlich lang erscheinen. Die Analogie beeindruckte mich dermaßen,

dass ich einen Begriff für unsere Erfahrung prägte: die »Illusion der Gültigkeit«.

Ich hatte meine erste kognitive Illusion entdeckt.

Jahrzehnte später erkenne ich viele der zentralen Themen meiner Forschungsarbeit – und dieses Buches – in dieser alten Geschichte. Unsere Erwartungen hinsichtlich der zukünftigen Leistungen der Soldaten waren ein eindeutiges Beispiel für Ersetzung und insbesondere für die Repräsentativitätsheuristik. Nachdem wir das Verhalten eines Soldaten in einer künstlichen Situation eine Stunde lang beobachtet hatten, glaubten wir zu wissen, wie gut er den Herausforderungen der Offiziersausbildung gewachsen wäre und ob er im Kampf eine Führungsrolle übernehmen würde. Unsere Vorhersagen berücksichtigten in keiner Weise die Regression zum Mittelwert – wir hatten keinerlei Vorbehalte dagegen, auf der Basis dürftiger Informationen zukünftiges Scheitern oder herausragenden Erfolg vorherzusagen. Dies war ein eindeutiger Fall von WYSIATI: Allein die verfügbare Information zählt! Wir hatten überzeugende Eindrücke von dem Verhalten, das wir beobachteten, und keine geeignete Methode, um unsere Unkenntnis der Faktoren, die letztlich bestimmen, welche Leistungen ein Kandidat als Offizier bringt, zu repräsentieren.

Im Rückblick ist das Bemerkenswerteste an der ganzen Geschichte die Tatsache, dass unsere Kenntnis der allgemeinen Regel – dass eine Vorhersage nicht möglich ist – sich nicht auf die Gewissheit unserer Überzeugungen in einzelnen Fällen auswirkte. Heute weiß ich, dass wir ganz ähnlich reagierten wie die Studenten von Nisbett und Borgida, als ihnen gesagt wurde, die meisten Menschen würden einem Fremden, der einen epileptischen Anfall erleidet, nicht helfen. Sie glaubten zweifellos den gezeigten Statistiken, aber die Basisraten beeinflussten nicht ihr Urteil darüber, ob eine Person, die sie im Video sahen, einem Fremden helfen würde oder nicht. Wie von Nisbett und Borgida gezeigt, widerstrebt es uns oftmals, das Besondere aus dem Allgemeinen abzuleiten.

Subjektives Vertrauen in ein Urteil ist keine wohldurchdachte Einschätzung der Wahrscheinlichkeit, mit der dieses Urteil zutreffend ist. Vertrauen ist ein Gefühl, in dem sich die Kohärenz der Informationen und die Leichtigkeit ihrer kognitiven Verarbeitung widerspiegeln. Es ist klug, Eingeständnisse der Ungewissheit ernst zu nehmen, aber Erklärungen hoher subjektiver Überzeugung sagen uns nur, dass eine Person in ihrem Kopf

eine kohärente Geschichte konstruiert hat, nicht unbedingt, dass diese Geschichte wahr ist.

## Der Irrglaube, einen guten Riecher für Aktien zu haben

Im Jahr 1984 besuchten Amos und ich mit unserem Freund Richard Thaler eine Wall-Street-Firma. Unser Gastgeber, ein hochrangiger Investmentmanager, hatte uns eingeladen, um mit uns über die Rolle von Urteilsfehlern bei der Kapitalanlage zu diskutieren. Ich kannte mich so wenig mit Finanzprodukten aus, dass ich nicht einmal wusste, was ich ihn fragen sollte, aber ich erinnere mich an einen Wortwechsel. »Wenn Sie eine Aktie verkaufen«, so fragte ich, »wer kauft sie dann?« Er antwortete mit einer vagen Handbewegung in Richtung des Fensters und gab damit zu verstehen, dass er davon ausging, der Käufer würde jemand sein, der große Ähnlichkeit mit ihm hatte. Das war seltsam: Was veranlasste eine Person dazu, gerade das zu kaufen, was jemand anders verkaufte? Was glaubten die Verkäufer zu wissen, was die Käufer nicht wussten?

Seit dieser Zeit haben sich meine Fragen über den Aktienmarkt zu einem größeren Rätsel verdichtet: Eine bedeutende Branche scheint weitgehend auf einer »Kompetenzillusion« *(illusion of skill)* zu basieren. Tagtäglich werden Milliarden von Aktien gehandelt, wobei viele Menschen Aktien eines bestimmten Unternehmens kaufen, die ihnen von anderen verkauft werden. Es ist nicht ungewöhnlich, dass an einem Tag 100 Millionen Aktien eines einzigen Unternehmens an der Börse den Besitzer wechseln. Die meisten Käufer und Verkäufer wissen, dass sie dieselben Informationen besitzen; sie tauschen die Aktien hauptsächlich deshalb, weil sie unterschiedlicher Meinung sind. Die Käufer glauben, die Aktien seien zu niedrig bewertet, und ihr Kurs werde wahrscheinlich steigen, während die Verkäufer die Aktien für überbewertet halten und glauben, ihr Kurs werde sinken. Die Frage ist, wieso Käufer und Verkäufer gleichermaßen der Ansicht sind, der gegenwärtige Preis sei unangemessen. Was lässt sie glauben, den angemessenen Preis besser als der Markt zu wissen? Für die meisten von ihnen ist diese Überzeugung eine Illusion.

Die Standardtheorie über die Funktionsweise des Aktienmarktes wird in ihren Grundzügen von allen Marktteilnehmern als richtig anerkannt. Alle Akteure im Kapitalanlage-Geschäft haben Burton Malkiels wunderbares Buch *Börsenerfolg ist kein Zufall* gelesen. Malkiel geht von der zentralen Annahme aus, dass der Kurs einer Aktie sämtliche verfügbaren

Informationen über den Wert des betreffenden Unternehmens und die besten Vorhersagen über dessen Zukunft widerspiegelt. Wenn einige Leute glauben, der Kurs einer Aktie werde morgen höher sein, werden sie heute mehr davon kaufen. Dies wiederum führt zu einem Kursanstieg. Wenn alle Vermögenswerte, die auf einem Markt gehandelt werden, den »richtigen« Preis besitzen, dann kann niemand erwarten, durch Kauf oder Verkauf einen Gewinn oder einen Verlust zu machen. Vollkommene Preise lassen keinen Raum für intelligente Kalküle, aber sie schützen den Leichtfertigen oder Unkundigen auch vor den Folgen seines Leichtsinns. Wir wissen heute jedoch, dass diese Theorie nicht stimmt. Viele Investoren machen bei ihren Wertpapiergeschäften durchgehend Verluste, eine Leistung, mit der es ein Dartpfeile werfender Schimpanse nicht aufnehmen könnte. Terry Odean, Professor für Finanzwirtschaft an der Universität von Kalifornien in Berkeley – ein ehemaliger Student von mir –, hat diese erstaunliche Tatsache als Erster nachgewiesen.[1]

Odean analysierte zunächst sämtliche Transaktionen, die von einzelnen Investoren über einen Zeitraum von sieben Jahren über 10 000 Brokerage-Konten (Abwicklungskonten für Wertpapierhandel) getätigt wurden. Er wertete jede Transaktion aus, die Investoren über diese Firma ausführten, insgesamt fast 163 000 Kauf- und Verkaufsgeschäfte. Diese große Datenmenge erlaubte es Odean, sämtliche Fälle zu identifizieren, in denen ein Anleger einen Teil seines Bestands an Anteilsscheinen eines bestimmten Wertes verkaufte und bald darauf eine andere Aktie kaufte. Der Anleger enthüllt durch diese Geschäfte, dass er eine eindeutige Erwartung bezüglich der zukünftigen Wertentwicklung der beiden Aktien hatte: Er erwartete, dass sich die Aktie, die er kaufte, besser entwickeln würde als die Aktie, die er verkaufte.

Um herauszufinden, ob diese Erwartungen begründet waren, verglich Terry Odean die Wertentwicklung der Aktie, die der Investor abgestoßen hatte, mit derjenigen, die er stattdessen gekauft hatte, und zwar über einen Zeitraum von einem Jahr nach Abschluss der Transaktion. Die Ergebnisse waren eindeutig negativ. Im Schnitt entwickelten sich die Aktien, die die Händler verkauften, besser als jene, die sie kauften, und zwar erheblich besser: um 3,2 Prozentpunkte pro Jahr, und dies unter Berücksichtigung der beachtlichen Transaktionskosten.

Man sollte sich daran erinnern, dass dies eine Aussage über Durchschnittswerte ist: Einige Anleger verzeichneten weit höhere

Wertzuwächse, andere weit niedrigere. Dennoch steht fest, dass es für die große Mehrheit der einzelnen Investoren besser gewesen wäre, wenn sie gar nichts unternommen hätten, als ihren Erwartungen zu folgen. Spätere Studien von Odean und seinem Kollege Brad Barber erhärteten diese Schlussfolgerung. In einem Aufsatz mit dem Titel »Wertpapiergeschäfte bedrohen Ihr Vermögen« zeigten sie, dass die aktivsten Händler die schlechtesten Ergebnisse hatten, während die Anleger, die am passivsten waren, die höchsten Renditen einstrichen. In einem anderen Aufsatz mit dem Titel »Jungen sind eben Jungen« wiesen sie nach, dass Männer ihre unbrauchbaren Ideen deutlich häufiger in die Tat umsetzten als Frauen und dass Frauen daher bessere Anlageergebnisse erzielen als Männer.[2]

Natürlich gibt es bei jeder Transaktion einen Geschäftspartner; dies sind in der Regel Finanzinstitute und professionelle Anleger, die bereit sind, die Fehler auszunutzen, die Privatanleger bei der Auswahl von abzustoßenden und zu kaufenden Aktie machen. Weitere Forschungen von Barber und Odean haben diese Fehler genauer beleuchtet. Privatanleger neigen dazu, ihre Gewinne zu realisieren, indem sie »Gewinner« verkaufen, also Aktien, die seit ihrem Kauf einen deutlichen Kursanstieg verzeichnet haben, während sie an ihren »Verlierern« festhalten.[3] Doch zu ihrem Pech zeigen aktuelle Gewinner kurzfristig eine bessere Wertentwicklung als aktuelle Verlierer, sodass die Anleger die falschen Aktien verkaufen. Und sie kaufen auch die falschen Aktien. Privatanleger wenden sich vorhersehbarerweise in Scharen Unternehmen zu, die ihre Aufmerksamkeit auf sich ziehen, weil sie Schlagzeilen machen. Professionelle Investoren reagieren selektiver auf Nachrichten.[4] Diese Befunde lassen den Ausdruck *smart money* (professionell verwaltete, rentable Anlagegelder), den Finanzprofis auf sich selbst beziehen, gerechtfertigt erscheinen.

Auch wenn Profis den Amateuren erhebliche Summen an Geld entlocken können, haben nur wenige Stockpicker (Investoren, die nach Einzeltiteln Ausschau halten, welche eine überdurchschnittliche Rendite versprechen) die erforderliche Sachkunde, um durchgängig, Jahr für Jahr, den Markt zu schlagen.[5] Professionelle Anleger inklusive Fondsmanager lassen ein elementares Zeichen für Kompetenz vermissen: anhaltenden Erfolg. Das Diagnosekriterium für Kompetenz ist die Konsistenz individueller Leistungsunterschiede. Die Logik ist einfach: Wenn individuelle Unterschiede in einem beliebigen Jahr gänzlich zufallsbedingt sind,

schwankt das Ranking der Investoren und Fonds in einer unberechenbaren Weise, und die Korrelation der Rangordnungen von Jahr zu Jahr ist gleich null. Wenn die Kompetenz der Anleger eine Rolle spielt, sind die Rankings stabiler. Die Beständigkeit individueller Unterschiede ist das Maß, mit dem wir das Vorhandensein von Kompetenz bei Golfern, Autohändlern, Kieferorthopäden oder flotten Kassierern in Mautstellen erfassen.

Investmentfonds werden von äußerst erfahrenen und tüchtigen Fachleuten gemanagt, die Aktien kaufen und verkaufen, um die bestmöglichen Ergebnisse für ihre Kunden herauszuholen. Dennoch ist die Datenlage nach über fünfzigjähriger Forschung eindeutig: Bei der großen Mehrheit der Fondsmanager gleicht die Auswahl von Einzeltiteln eher einem Würfel- als einem Pokerspiel. Im Allgemeinen ist die Wertentwicklung bei zwei von drei Investmentfonds in jedem beliebigen Jahr schlechter als die des Gesamtmarktes.[6]

Noch wichtiger aber ist, dass die Korrelation zwischen aufeinanderfolgenden jährlichen Anlageergebnissen von Investmentfonds sehr klein ist, kaum größer als null. Die erfolgreichen Fonds in jedem beliebigen Jahr haben einfach das meiste Glück; sie haben gut gewürfelt. Es besteht allgemeine Einigkeit unter Forschern, dass fast alle Stockpicker, ob sie es wissen oder nicht – und nur wenige wissen es –, ein Zufallsspiel spielen. Wertpapierhändler haben subjektiv das Gefühl, in einer Situation großer Ungewissheit rationale, wohlbegründete Vermutungen anzustellen. In hocheffizienten Märkten aber sind auch fundierte Vermutungen nicht treffgenauer als blinde Vermutungen.

Vor einigen Jahren hatte ich eine ungewöhnliche Gelegenheit, die Illusion, dass finanzwirtschaftliche Sachkunde bessere Anlageergebnisse gewährleistet, genauer zu prüfen. Ich war eingeladen worden, in einem Unternehmen, das sehr reichen Kunden Vermögensberatung und andere Dienstleistungen anbietet, einen Vortrag vor einer Gruppe von Anlageberatern zu halten. Ich bat im Vorfeld um einige Daten zur Vorbereitung meiner Präsentation und erhielt einen kleinen Schatz: eine Tabelle, in der die Anlageergebnisse von etwa 25 anonymen Vermögensberatern für jeweils acht aufeinanderfolgende Jahre aufgelistet waren. Der Punktwert jedes Beraters für jedes Jahr war der bestimmende Faktor für die Höhe seiner Jahresprämie. Es war leicht, die Berater nach ihrer jeweiligen jährlichen Leistung einzustufen und herauszufinden, ob es

beständige Erfolgsunterschiede zwischen ihnen gab und ob dieselben Berater anhaltend, Jahr für Jahr, bessere Ergebnisse für ihre Kunden herausholten.[7]

Um diese Frage zu beantworten, berechnete ich Korrelationskoeffizienten zwischen den Rangplätzen in jedem Paar von Jahren: Jahr eins mit Jahr zwei, Jahr eins mit Jahr drei und so weiter, bis zu Jahr sieben mit Jahr acht. Das ergab 28 Korrelationskoeffizienten, einen für jedes Paar von Jahren. Ich kannte die Theorie und rechnete damit, leichte Belege für beständige kompetenzbasierte Erfolgsunterschiede zu finden. Doch überraschenderweise betrug der Mittelwert der 28 Korrelation 0,1. Anders gesagt: null. Die konstanten Korrelationen, die auf Kompetenzunterschiede hingedeutet hätten, waren nicht zu finden. Die Ergebnisse glichen dem, was man bei einem Würfelspiel, nicht bei einem Geschicklichkeitsspiel erwarten würde.

Niemand in dem Unternehmen schien sich der Eigenart des Spiels, das seine Stockpicker spielten, bewusst zu sein. Die Berater selbst sahen sich als kompetente Fachleute, die einer seriösen Arbeit nachgingen, und ihre Vorgesetzten sahen das genauso. Am Vorabend des Seminars aßen Richard Thaler und ich mit einigen der Topmanager der Firma zu Abend, jenen Personen, die über die Höhe der Prämie entscheiden. Wir baten sie, die Korrelation zwischen den jährlichen Rangplätzen einzelner Berater zu schätzen. Sie glaubten zu wissen, was nun kommen würde, und sagten lächelnd »Nicht sehr hoch« oder »Die Leistung schwankt bestimmt«. Es stellte sich jedoch schnell heraus, dass niemand erwartet hätte, dass die mittlere Korrelation gleich null ist.

Unsere Botschaft an die Führungskräfte lautete, dass die Firma, zumindest was die Zusammenstellung der Wertpapierdepots betraf, bloßes Glück so belohnte, als wäre es Können. Diese Neuigkeit hätte sie eigentlich erschüttern müssen, aber das tat sie nicht. Es gab keinen Anhaltspunkt dafür, dass sie uns nicht glaubten. Wie auch? Schließlich hatten wir ihre eigenen Daten analysiert, und sie waren intelligent genug, um selbst die Schlussfolgerungen daraus zu ziehen, die wir aus Höflichkeit nicht laut aussprachen. Wir speisten in aller Seelenruhe zu Ende, und ich hatte nicht den geringsten Zweifel daran, dass unsere Ergebnisse und die Folgerungen, die sich aufdrängten, schnell unter den Teppich gekehrt würden und dass in der Firma alles so weiterginge wie zuvor. Die Kompetenzillusion ist nicht nur ein individueller Urteilsfehler; sie ist tief in der Kultur der Wirtschaft

verwurzelt. Tatsachen, die Grundannahmen infrage stellen – und dadurch das Auskommen und die Selbstachtung von Menschen bedrohen –, werden einfach ausgeblendet. Unsere Psyche verarbeitet sie nicht. Dies gilt ganz besonders für statistische Studien über die Leistungsfähigkeit, die Basisraten-Informationen liefern, welche von Menschen im Allgemeinen ignoriert werden, wenn sie ihren Eindrücken aus persönlichen Erfahrungen zuwiderlaufen.

Am nächsten Morgen eröffneten wir den Beratern unsere Ergebnisse, und sie reagierten genauso zurückhaltend. Ihre persönliche Erfahrung mit der umsichtigen Beurteilung komplexer Probleme war für sie viel überzeugender als eine undurchsichtige statistische Tatsache. Nach Abschluss des Seminars fuhr mich einer der Manager, mit denen ich am Vorabend gespeist hatte, zum Flughafen. Er sagte mir, mit einer Spur von Gereiztheit: »Ich habe sehr viel für diese Firma geleistet, und das kann mir niemand nehmen.« Ich lächelte und sagte nichts. Aber ich dachte: »Nun, ich habe es dir heute Morgen genommen. Wenn dein Erfolg sich größtenteils dem Zufall verdankt, wie viel Anerkennung hast du dann dafür verdient?«

## Kompetenz und Gültigkeit – warum wir diesen Illusionen unterliegen

Kognitive Illusionen können hartnäckiger sein als visuelle Illusionen. Was Sie über die Müller-Lyer-Illusion lernten, hat nicht dazu geführt, dass Sie die Linien jetzt anders wahrnehmen, aber es hat Ihr Verhalten verändert. Sie wissen jetzt, dass Sie Ihrem Eindruck von der Länge der Linie mit angefügten »Schwanzflossen« nicht trauen können, und Sie wissen auch, dass Sie bei der gängigen Darstellung der Müller-Lyer-Illusion dem, was Sie sehen, nicht trauen können. Wenn man Sie nach der Länge der Linien fragt, werden Sie Ihre sachlich fundierte Überzeugung äußern, nicht die optische Täuschung, die Sie nach wie vor sehen. Als meine Kollegen und ich dagegen in der Armee zur Kenntnis nehmen mussten, dass unsere Tests zur Bewertung der Führungsfähigkeit nur eine geringe prognostische Gültigkeit besaßen, fanden wir uns intellektuell mit dieser Tatsache ab, aber sie hatte keinen Einfluss auf unsere Gefühle oder unsere anschließenden Handlungen. Die Reaktion, auf die wir in der Vermögensberatungsgesellschaft stießen, war noch extremer. Ich bin fest davon überzeugt, dass die Botschaft, die Thaler und ich den Führungskräften und den Vermögensverwaltern zukommen ließen, sofort in einem dunklen Winkel des Gedächtnisses abgelegt wurde, in dem sie keinen Schaden anrichten würde.

Wieso halten sowohl Amateur- als auch professionelle Anleger hartnäckig an der Überzeugung fest, dass sie den Markt schlagen können, obwohl dies der Wirtschaftstheorie widerspricht, die die meisten von ihnen akzeptieren, und auch dem, was sie aus einer nüchternen Bewertung ihrer persönlichen Erfahrungen lernen könnten? Viele Themen der vorangehenden Kapitel begegnen uns wieder, wenn wir die Häufigkeit und Beständigkeit der Kompetenzillusion im Finanzdienstleistungssektor erklären wollen.

Die stärkste psychische Ursache der Illusion ist zweifellos die Tatsache, dass Menschen, die Stock-picking betreiben, hohe kognitive Leistungen erbringen. Sie konsultieren wirtschaftliche Daten und Vorhersagen, sie analysieren Gewinn- und Verlustrechnungen sowie Bilanzen, sie bewerten die Qualität der Unternehmensführung, und sie schätzen die Konkurrenz

ein. All dies ist fachlich anspruchsvolle Arbeit, die eine umfassende Ausbildung erfordert, und Menschen, die diese Arbeit erledigen, haben die unmittelbare (und zutreffende) Erfahrung, diese Fähigkeiten einzusetzen. Leider ist die Fähigkeit, die Geschäftsaussichten eines Unternehmens sachgerecht zu beurteilen, nicht ausreichend für erfolgreiche Aktiengeschäfte, bei denen die Schlüsselfrage lautet, ob alle relevanten Informationen über eine Unternehmung bereits in dem Kurs ihrer Aktie berücksichtigt wurden. Wertpapierhändlern fehlt offenbar die Fähigkeit, diese entscheidende Frage zu beantworten, aber sie scheinen ihre Unwissenheit zu ignorieren. So, wie ich es bei mir selbst bemerkt hatte, als ich Kadetten auf dem Hindernisfeld beobachtete, ist subjektives Überzeugtsein ein Gefühl, kein Urteil – das gilt auch für Händler. Unser Verständnis kognitiver Leichtigkeit und assoziativer Kohärenz verortet dieses Gefühl der subjektiven Überzeugung fest in System 1.

Schließlich werden die Illusionen der Gültigkeit und Kompetenz von einer mächtigen Berufskultur gestützt. Wir wissen, dass Menschen von einem unerschütterlichen Glauben an eine Überzeugung, und sei sie noch so absurd, erfüllt sein können, wenn sie darin von einer Gruppe Gleichgesinnter bestärkt werden. Angesichts der Berufskultur im Finanzdienstleistungssektor ist es nicht verwunderlich, dass sehr viele Personen in diesem Sektor sich zu den wenigen Auserwählten zählen, die ihres Erachtens etwas können, was andere nicht können.

## Die Illusionen von Experten

Die Annahme, die Zukunft lasse sich nicht vorhersagen, wird Tag für Tag durch die Leichtigkeit untergraben, mit der die Vergangenheit erklärt wird. Wie Nassim Taleb in *Der Schwarze Schwan* ausführt, fällt es uns deshalb schwer, die Grenzen unserer Vorhersagefähigkeit anzuerkennen, weil wir dazu neigen, kohärente Erzählungen über vergangene Ereignisse zu konstruieren und daran zu glauben. Im Rückblick ergibt alles einen Sinn, eine Tatsache, die Finanzexperten jeden Abend ausnutzen, wenn sie überzeugende Erklärungen für die Ereignisse des Tages liefern. Und wir können nicht die mächtige Intuition unterdrücken, dass das, was heute im Rückblick einen Sinn zu ergeben scheint, gestern vorhersagbar war. Die Illusion, wir verstünden die Vergangenheit, fördert die Überschätzung unserer Fähigkeit, die Zukunft vorherzusagen.

Das oft verwendete Bild vom »Marsch der Geschichte« legt Ordnung und eine eindeutige Richtung nahe. Anders als ein Bummel oder ein Spaziergang führen Märsche nicht »ins Blaue hinein«. Wir glauben, dass wir die Vergangenheit dadurch erklären können, dass wir uns entweder auf soziale Massenbewegungen und kulturelle sowie technologische Entwicklungen oder auf die Absichten und Fähigkeiten einiger weniger bedeutender Männer konzentrieren. Die These, dass sich bedeutende historische Ereignisse rein dem Zufall verdanken, ist zutiefst erschütternd, aber sie ist nachweislich wahr. Es ist schwer, die Geschichte des 20. Jahrhunderts einschließlich seiner sozialen Massenbewegungen zu verstehen, ohne auf die Rolle von Hitler, Stalin und Mao Tsetung einzugehen. Aber es gab einen bestimmten Zeitpunkt, unmittelbar vor der Befruchtung eines Eies, als eine 50-prozentige Chance bestand, dass der Embryo, der zu Hitler wurde, ein weibliches Geschlecht erhält. Nimmt man die drei Ereignisse zusammen, bestand eine Wahrscheinlichkeit von 1:8 eines 20. Jahrhunderts ohne irgendeinen der drei Erzschurken, und es lässt sich unmöglich behaupten, dass die Geschichte ohne sie annähernd den gleichen Verlauf genommen hätte. Die Befruchtung dieser drei Eier hatte weitreichende Folgen, und sie führt die Annahme, langfristige Entwicklungen seien vorhersagbar, ad absurdum.

Trotzdem hält sich die Illusion gültiger Vorhersagen, was von Menschen ausgenutzt wird, deren Geschäft Prognosen sind – nicht nur Finanzexperten, sondern auch andere Fachleute in Wirtschaft und Politik. Fernseh- und Rundfunksender sowie Zeitungen haben ihre Expertengremien, deren Aufgabe es ist, die jüngsten Ereignisse zu kommentieren und die Zukunft vorherzusagen. Zuschauer und Leser haben den Eindruck, dass sie Informationen erhalten, die irgendwie privilegiert oder zumindest äußerst aufschlussreich sind. Und es besteht kein Zweifel daran, dass die Experten und ihre Förderer ernsthaft glauben, solche Informationen zu liefern. Der Psychologe Philip Tetlock von der Universität von Pennsylvania erklärte diese sogenannten Expertenprognosen in einer bahnbrechenden Studie, die er in seinem 2005 erschienenen Buch *Expert Political Judgment: How Good Is It?* veröffentlichte. Tetlock hat den Standard für jede zukünftige Diskussion dieses Themas definiert.

Tetlock befragte 284 Personen, die ihren Lebensunterhalt als »Kommentatoren oder Berater für politische und ökonomische Trends« verdienen. Er bat sie, die Wahrscheinlichkeiten zu beurteilen, dass bestimmte Ereignisse in nicht allzu ferner Zukunft eintreten, und zwar sowohl in Regionen der Welt, auf die sie sich spezialisiert hatten, als auch in Regionen, über die sie weniger gut Bescheid wussten. Würde Gorbatschow durch einen Staatsstreich gestürzt? Würden die Vereinigten Staaten am Persischen Golf in den Krieg ziehen? Welches Land würde das nächste große Schwellenland werden? Insgesamt trug Tetlock über 80 000 Vorhersagen zusammen. Er fragte die Experten auch, wie sie zu ihren Schlussfolgerungen gelangten, wie sie reagierten, wenn sie widerlegt wurden, und wie sie Informationen bewerteten, die ihren Standpunkt nicht untermauerten. Die Befragten sollten die Wahrscheinlichkeiten dreier alternativer Ergebnisse in jedem Fall einschätzen: das Fortdauern des Status quo, mehr von einer Sache wie politische Freiheit oder Wirtschaftswachstum oder weniger davon.

Die Ergebnisse waren niederschmetternd. Die Experten zeigten eine schlechtere Leistung, als wenn sie einfach alle drei möglichen Ergebnisse mit der gleichen Wahrscheinlichkeit eingestuft hätten. Anders gesagt: Menschen, die ihre Zeit damit verbringen – und ihren Lebensunterhalt damit verdienen –, sich gründlich mit einem bestimmten Sachgebiet zu beschäftigen, erstellen schlechtere Vorhersagen als Dartpfeile werfende Affen, die ihre »Entscheidungen« gleichmäßig über alle Optionen verteilt

hätten. Selbst auf dem Gebiet, das sie am besten kannten, waren Experten nicht deutlich besser als Nichtexperten.

Diejenigen, die mehr wissen, liefern geringfügig bessere Vorhersagen als diejenigen, die weniger wissen. Aber diejenigen mit dem meisten Wissen sind oftmals weniger zuverlässig. Das ist darauf zurückzuführen, dass jemand, der mehr Wissen erwirbt, eine verstärkte Illusion von seinen Fähigkeiten entwickelt und diese in einer unrealistischen Weise überschätzt. »Wir erreichen beunruhigend schnell den Punkt eines abnehmenden prädiktiven Grenzertrags von Wissen«, schreibt Tetlock. »In diesem Zeitalter der akademischen Hyperspezialisierung gibt es keinen Grund, anzunehmen, dass Personen, die Beiträge in Top-Zeitschriften publizieren – bekannte Politikwissenschaftler, Regionalwissenschaftler, Ökonomen usw. – neu auftretende Situationen besser ›lesen‹ können als Journalisten oder aufmerksame Leser der *New York Times*.«[8] Je berühmter der Prognostiker, umso extravaganter waren seine Prognosen, fand Tetlock heraus. »Experten, die hoch im Kurs standen«, schreibt er, »überschätzten sich selbst stärker als ihre Kollegen, die fern des Rampenlichts ein kümmerliches Dasein fristen.«

Tetlock fand auch heraus, dass Experten nur widerwillig zugaben, sich geirrt zu haben, und wenn sie gezwungen waren, einen Fehler zuzugeben, hatten sie jede Menge Ausreden parat: Sie hätten sich nur im Zeitpunkt geirrt, ein unvorhersehbares Ereignis sei dazwischengekommen, oder sie hätten sich zwar geirrt, aber aus den richtigen Gründen. Experten sind schließlich auch nur Menschen. Sie sind geblendet von ihrer Brillanz und hassen es, danebenzuliegen. Nicht ihre Überzeugungen, sondern ihre Art zu denken, führe sie in die Irre, so Tetlock. Er verwendet die Begriffe, die Isaiah Berlin in seinem Tolstoi-Essay »Der Igel und der Fuchs« einführte: Igel »haben den großen Durchblick« und verfügen über eine umfassende Theorie; sie erklären konkrete Ereignisse innerhalb eines kohärenten Bezugssystems, reagieren mit Unverständnis auf diejenigen, die ihre Sichtweise nicht teilen, und sind überzeugt von der Richtigkeit ihrer Vorhersagen. Fehler räumen sie nur sehr widerwillig ein. Für Igel ist eine falsche Prognose fast immer nur »zeitlich verfehlt« oder »fast richtig«. Sie sind von sich selbst überzeugt und ihrer Sache sicher, und genau das wünschen sich Sendeleiter in ihren Fernsehsendungen. Zwei Igel, die in einer Frage gegensätzliche Standpunkte vertreten, sorgen für eine unterhaltsame Show, in der jeder die idiotischen Ideen des Gegners angreift.

Füchse hingegen denken komplex. Sie glauben nicht, dass ein großer Faktor den Marsch der Geschichte antreibt (so werden sie kaum der Auffassung zustimmen, Ronald Reagan habe im Alleingang den Kalten Krieg beendet, indem er der Sowjetunion die Stirn geboten habe). Vielmehr sind die Füchse davon überzeugt, dass die Wirklichkeit das Produkt der Wechselwirkungen vieler verschiedener Agenten und Kräfte ist, einschließlich des blinden Zufalls, die oftmals zu unvorhersehbaren Ergebnissen mit weitreichenden Folgen führen. In Tetlocks Studie schnitten die Füchse besser ab, auch wenn ihre Leistung insgesamt sehr schlecht war. Sie werden seltener als Igel zu Fernsehdiskussionen eingeladen.

## Die Experten können nichts dafür – die Welt ist eben kompliziert

In diesem Kapitel ging es nicht darum, zu verdeutlichen, dass Menschen, die Prognosen erstellen, viele Fehler machen; das versteht sich von selbst. Die erste Lektion lautet, dass Vorhersagefehler unvermeidlich sind, weil die Welt nicht vorhersagbar ist. Die zweite lautet, dass starkes subjektives Überzeugtsein kein zuverlässiger Indikator für Richtigkeit ist (eine schwache Überzeugung könnte informativer sein).

Kurzfristige Trends lassen sich vorhersagen, und das Verhalten sowie die Leistungsfähigkeit lassen sich ziemlich genau aus dem früheren Verhalten und vorhergehenden Leistungen prognostizieren. Aber wir sollten nicht erwarten, dass sich die Leistung bei der Offiziersausbildung und im Kampfeinsatz aus dem Verhalten auf einem Hindernisfeld vorhersagen lässt, denn das Verhalten beim Test und in der realen Welt wird von vielen Faktoren bestimmt, die spezifisch für die konkrete Situation sind. Man entferne eine Person mit starkem Durchsetzungsvermögen aus einer Gruppe von acht Kandidaten, und die Persönlichkeiten aller anderen werden sich anscheinend verändern. Wenn ein Offiziersanwärter beim Scharfschützentest ein paar Zentimeter danebenschießt, kann dies das Aus für ihn bedeuten. Ich behaupte nicht, dass alle Tests unbrauchbar sind – wenn ein Test ein wichtiges Ergebnis mit einer Zuverlässigkeit von 0,20 oder 0,30 prognostiziert, sollte man ihn anwenden. Aber man sollte nicht mehr erwarten. Man sollte wenig oder gar nichts von Stockpickern an der Wall Street erwarten, die hoffen, zukünftige Kursentwicklungen besser vorhersagen zu können als der Markt. Und man sollte nicht viel von Experten erwarten, die langfristige Prognosen erstellen – auch wenn sie Nützliches über die nahe Zukunft wissen mögen. Die Linie, die die möglicherweise vorhersagbare Zukunft von der unvorhersagbar fernen Zukunft trennt, muss erst noch gezogen werden.

### Zum Thema »Illusorische Gewissheit«

»Er weiß, dass die Daten darauf hindeuten, dass sich seine Krankheit in einer Weise entwickeln wird, die kaum vorhersagbar ist. Wie kann er in diesem Fall so optimistisch sein? Hört sich nach trügerischer Gewissheit an.«

»Sie hat alles, was sie weiß, zu einer kohärenten Geschichte verarbeitet, und diese Kohärenz verschafft ihr ein tiefes Wohlgefühl.«

»Was veranlasst ihn, zu glauben, er wäre intelligenter als der Markt? Ist das eine Kompetenzillusion?«

»Sie ist ein Igel. Sie verfügt über eine Theorie, die alles erklärt, und diese vermittelt ihr die Illusion, die Welt zu verstehen.«

»Die Frage ist nicht, ob diese Experten gut ausgebildet sind. Die Frage ist, ob ihre Welt vorhersagbar ist.«

# *21. Intuitionen und Formeln*

Paul Meehl war ein eigenartiger und wunderbarer Mensch und einer der vielseitigsten Psychologen des 20. Jahrhunderts. Er hatte Lehraufträge für so unterschiedliche Gebiete wie Psychologie, Jura, Psychiatrie, Neurologie und Philosophie. Er publizierte auch über Religion, Politikwissenschaft und das Lernverhalten von Ratten. Meehl war nicht nur ein Forscher mit profunden Statistik-Kenntnissen und ein erbitterter Kritiker nichtssagender Behauptungen in der klinischen Psychologie – er war auch praktizierender Psychoanalytiker. Er schrieb gedankenreiche Aufsätze über die philosophischen Grundlagen psychologischer Forschung, die ich als Student fast auswendig lernte. Ich habe Meehl nie kennengelernt, aber nachdem ich sein Buch *Clinical vs. Statistical Prediction: A Theoretical Analysis and a Review of the Evidence* gelesen hatte, war er einer meiner Helden.

In dem schmalen Band, den er später »mein verstörendes kleines Buch« nannte, wertete Meehl die Ergebnisse von zwanzig Studien aus, in denen analysiert wurde, ob *klinische* Vorhersagen auf der Basis subjektiver Eindrücke von Fachleuten zutreffender waren als *statistische* Vorhersagen, die man erhielt, wenn man einige wenige Scores (Punktwerte) oder Ratings nach einer bestimmten Regel kombinierte. In einer typischen Untersuchung sagten geschulte Studienberater die Noten von Erstsemestern am Ende des Studienjahres vorher. Die Berater interviewten jeden Studenten 45 Minuten lang. Sie hatten auch Zugriff auf die Highschool-Noten, die Ergebnisse mehrerer Eignungstests und eine vierseitige persönliche Stellungnahme. Der statistische Algorithmus benutzte nur einen Bruchteil dieser Informationen: Highschool-Noten und einen Eignungstest. Trotzdem lieferte die Formel zutreffendere Ergebnisse als 11 der 14 Berater. Meehl berichtete über ganz ähnliche Resultate bei einer Vielzahl anderer Prognoseergebnisse wie etwa der Verletzung von Bewährungsauflagen, dem erfolgreichen Abschluss einer Pilotenausbildung und der Rückfälligkeit von Straftätern.

Wie nicht anders zu erwarten, rief Meehls Buch Empörung und Ungläubigkeit bei klinischen Psychologen hervor, und die Kontroverse, die

es auslöste, hat einen Strom von Forschungsarbeiten hervorgebracht, der auch heute noch, fünfzig Jahre nach seiner Veröffentlichung, nicht versiegt ist. Die Anzahl der Studien, in denen klinische und statistische Vorhersagen verglichen wurden, ist auf ungefähr 200 gestiegen, aber der Stand in dem Wettkampf zwischen Algorithmen und Menschen hat sich nicht verändert. Bei etwa 60 Prozent der Studien erwiesen sich die Algorithmen als erheblich treffgenauer. Die anderen Vergleiche ergaben ein Unentschieden, aber ein Unentschieden ist gleichbedeutend mit einem Sieg für die statistischen Regeln, die im Allgemeinen viel kostengünstiger sind als das Urteil von Experten. Keine Ausnahme wurde glaubhaft dokumentiert.

Das Spektrum vorhergesagter Ergebnisse erstreckt sich mittlerweile sogar auf medizinische Variablen wie die Überlebenszeit von Krebspatienten, die Länge von Krankenhausaufenthalten, die Diagnose von Herzkrankheiten und die Anfälligkeit von Neugeborenen für den plötzlichen Kindstod; wirtschaftliche Maßnahmen wie etwa die Erfolgsaussichten neu gegründeter Unternehmen, die Bewertung von Kreditrisiken durch Banken und die zukünftige Zufriedenheit von Arbeitnehmern am Arbeitsplatz; Fragen, die für Behörden von Interesse sind, wie etwa die Beurteilung der Eignung von Pflegeeltern, das Rückfallrisiko bei jugendlichen Straftätern und die Wahrscheinlichkeit anderer Formen gewalttätigen Verhaltens; und diverse Ergebnisse wie etwa die Beurteilung wissenschaftlicher Vorträge, die Sieger bei Footballspielen und die zukünftigen Preise von Bordeauxweinen. Jeder dieser Tatbestände ist mit einem erheblichen Maß an Ungewissheit und Unvorhersagbarkeit verbunden. Sie werden »Umgebungen niedriger Validität« genannt. In allen Fällen erbrachte ein einfacher Algorithmus mindestens ebenso genaue Vorhersagen wie Experten.

Dreißig Jahre nach der Veröffentlichung seines Buches erklärte Meehl mit berechtigtem Stolz: »Es gibt keine andere Kontroverse in den Sozialwissenschaften, bei der die Ergebnisse einer so großen Zahl qualitativ unterschiedlicher Studien so einheitlich in dieselbe Richtung weisen wie bei dieser.«[1]

Der an der Universität Princeton lehrende Ökonom und Weinliebhaber Orley Ashenfelter hat anhand eines konkreten Fallbeispiels schlüssig nachgewiesen, dass einfache statistische Berechnungen weltberühmte Experten übertreffen. Ashenfelter wollte den zukünftigen Wert erlesener Bordeauxweine auf der Basis von Informationen vorhersagen, die in ihrem

Herstellungsjahr verfügbar waren. Die Frage ist wichtig, weil Spitzenweine erst nach jahrelanger Reifung ihre hervorragende Qualität erreichen, und die Preise ausgereifter Weine vom selben Weingut schwanken erheblich zwischen verschiedenen Jahrgängen; der Wert von Flaschen, die im Abstand von nur zwölf Monaten abgefüllt wurden, kann sich um das Zehnfache oder noch mehr unterscheiden.[2] Die Fähigkeit, zukünftige Preise vorherzusagen, ist äußerst nützlich, weil Investoren Weine in der Erwartung kaufen, dass ihr Wert steigen wird – ganz ähnlich wie bei Kunstwerken.

Es besteht allgemeines Einvernehmen darüber, dass die unterschiedliche Jahrgangsgüte nur mit Wetterschwankungen während der Traubenwachstumsperiode zusammenhängen kann. Die besten Weine werden produziert, wenn der Sommer warm und trocken ist, was die Produzenten von Bordeauxweinen vermutlich zu Nutznießern der globalen Erwärmung machen wird. Auch ein feuchtes Frühjahr hilft der Branche, denn es steigert die Erntemenge, ohne sich in nennenswertem Umfang auf die Weinqualität auszuwirken. Ashenfelter wandelte das herkömmliche Wissen in eine statistische Formel um, die den Preis eines Weines – von einem bestimmten Weingut und eines bestimmten Jahrgangs – mithilfe von drei Wettervariablen vorhersagt: der Durchschnittstemperatur während der sommerlichen Wachstumsperiode, der Niederschlagsmenge zur Erntezeit und des Gesamtniederschlags im vorangegangenen Winter. Seine Formel liefert genaue Preisvorhersagen über Jahre und sogar Jahrzehnte in die Zukunft hinein. Tatsächlich sagt seine Formel zukünftige Preise sehr viel genauer vorher, als es die aktuellen Preise junger Weine tun. Dieses neue Beispiel eines »Meehl-Musters« stellt die Fähigkeiten der Experten infrage, deren Einschätzungen den anfänglichen Preis maßgeblich beeinflussen. Es stellt auch die ökonomische Standardtheorie infrage, der zufolge sich in den Preisen alle verfügbaren Informationen einschließlich des Wetters widerspiegeln sollten. Ashenfelters Formel liefert äußerst genaue Ergebnisse – die Korrelation zwischen seinen Vorhersagen und den tatsächlichen Preisen liegt bei über 0,90.

Weshalb sind Experten Algorithmen unterlegen? Meehl vermutete, ein Grund liege darin, dass Experten besonders clever sein wollen, unkonventionell denken und bei ihren Vorhersagen komplexe Kombinationen von Merkmalen berücksichtigen. Komplexität mag in ungewöhnlichen Fällen helfen, aber sie verringert im Allgemeinen eher die Prognosegenauigkeit. Einfache Merkmalskombinationen sind besser.

Etliche Studien haben gezeigt, dass menschliche Entscheider einer Vorhersageformel unterlegen sind, selbst wenn man ihnen das Ergebnis der Formel mitteilt! Sie glauben, sie könnten sich über die Formel hinwegsetzen, weil sie zusätzliche Informationen über den Fall besitzen, aber damit liegen sie meistens falsch. Laut Meehl gibt es nur wenige Umstände, unter denen es sich empfiehlt, eine Formel durch ein Urteil zu ersetzen. In einem berühmten Gedankenexperiment beschrieb er eine Formel, die vorhersagt, ob eine bestimmte Person heute Abend ins Kino geht. Er wies darauf hin, es sei angemessen, die Formel zu ignorieren, wenn man die Information erhalte, dass sich die betreffende Person heute ein Bein gebrochen habe. Der Name »Beinbruch-Regel« hat sich durchgesetzt. Der springende Punkt ist der, dass Beinbrüche sehr selten – aber auch entscheidend – sind.

Eine weitere Ursache für die Unterlegenheit von Expertenurteilen liegt darin, dass Menschen bei summarischen Urteilen auf der Grundlage komplexer Informationen in unverbesserlicher Weise inkonsistent sind. Gebeten, dieselben Informationen zweimal zu beurteilen, geben sie oftmals unterschiedliche Antworten. Das Ausmaß der Inkonsistenz gibt häufig Anlass zu begründeter Sorge. Erfahrene Radiologen, die Röntgenaufnahmen der Brust als »unauffällig« oder »auffällig« einstufen sollen, widersprechen sich in 20 Prozent der Fälle, wenn ihnen dasselbe Bild bei verschiedenen Gelegenheiten gezeigt wird.[3] Eine Studie über 101 unabhängige Wirtschaftsprüfer, die die Zuverlässigkeit interner Revisionen beurteilen sollten, enthüllte ein ähnliches Maß an Inkonsistenz.[4] Eine Übersicht über 41 eigenständige Studien über die Zuverlässigkeit der Urteile von Wirtschaftsprüfern, Pathologen, Psychologen, Führungskräften und anderen Fachkräften deutet darauf hin, dass dieser Grad der Inkonsistenz typisch ist, auch wenn ein Fall innerhalb weniger Minuten erneut beurteilt wird.[5] Unzuverlässige Urteile erlauben keine gültigen Vorhersagen.

Die weitverbreitete Inkonsistenz ist vermutlich auf die extreme Kontextabhängigkeit von System 1 zurückzuführen. Wir wissen aus Priming-Studien, dass unbemerkte Stimuli in unserer Umgebung einen erheblichen Einfluss auf unser Denken und Handeln haben. Diese Einflüsse schwanken von einem Moment zum nächsten. Die kurze Annehmlichkeit einer kühlen Brise an einem heißen Tag mag dazu führen, dass man zu einer

etwas positiveren und optimistischeren Einschätzung dessen gelangt, was man gerade beurteilt. Die Chancen für einen Straftäter, auf Bewährung freigelassen zu werden, mögen sich in der Zeit, die zwischen aufeinanderfolgenden Esspausen im Terminplan des Bewährungsrichters vergeht, stark verändern.[6] Weil wir nur wenig direktes Wissen darüber besitzen, was in unserem Kopf geschieht, werden wir nie wissen, dass wir unter ganz geringfügig anderen Bedingungen ein anderes Urteil oder eine andere Entscheidung getroffen hätten. Formeln leiden nicht an solchen Problemen. Wenn sie denselben Input erhalten, stoßen sie immer die gleiche Antwort aus. Wenn die Vorhersagbarkeit gering ist – wie es bei den meisten der von Meehl und seinen Nachfolgern ausgewerteten Studien der Fall war –, zerstört Inkonsistenz jegliche prognostische Gültigkeit.

Die Forschungsergebnisse legen eine überraschende Schlussfolgerung nahe: Um die Vorhersagegenauigkeit zu maximieren, sollten abschließende Entscheidungen Formeln überlassen werden, insbesondere in Umgebungen mit geringer Validität. Bei der Zulassung zum Medizinstudium zum Beispiel wird die letzte Entscheidung oftmals von den Mitgliedern des Lehrkörpers getroffen, die den Bewerber interviewen. Die Daten sind fragmentarisch, aber es gibt gute Gründe für eine Vermutung: Das Führen eines Bewerbungsgesprächs wird vermutlich die Genauigkeit des Auswahlverfahrens verringern, wenn die Interviewer auch die endgültige Zulassungsentscheidung treffen. Weil Interviewer ihre Intuitionen überschätzen, messen sie ihren persönlichen Eindrücken zu viel Gewicht bei, während sie den Stellenwert anderer Informationsquellen unterschätzen, was die Gültigkeit ihrer Urteile verringert.[7] In ähnlicher Weise haben die Experten, die die Qualität unreifen Weins beurteilen, um seinen zukünftigen Preis vorherzusagen, eine Informationsquelle, die ihr Urteil eher verschlechtern als verbessern wird: Sie können den Wein probieren. Und selbst wenn sie ganz genau wissen, wie sich das Wetter auf die Weinqualität auswirkt, können sie es nicht mit der Konsistenz einer Formel aufnehmen.

Die wichtigste Entwicklung auf diesem Gebiet seit Meehls ursprünglicher Arbeit ist Robyn Dawes' berühmter Aufsatz »The Robust Beauty of Improper Linear Models in Decision Making«.[8] Die gängige statistische Vorgehensweise in den Sozialwissenschaften besteht darin, verschiedene

Prädiktoren (Vorhersagevariablen) zu gewichten, indem man einem »multiple Regression« genannten Algorithmus folgt, der heute in konventionellen Computerprogrammen enthalten ist. Die Logik der multiplen Regression ist unanfechtbar: Sie findet die optimale Formel, um eine gewichtete Kombination der Prädiktoren zusammenzustellen. Dawes wies jedoch darauf hin, dass der komplexe statistische Algorithmus keinen beziehungsweise keinen nennenswerten Zusatznutzen bringt. Ebenso gut kann man eine Reihe von Scores auswählen, die eine gewisse Prognosegültigkeit besitzen, und die Werte dann so anpassen, dass sie vergleichbar werden (indem man Standardwerte oder -rangzahlen verwendet). Eine Formel, die diese gleich gewichteten Prädiktoren kombiniert, sagt neue Fälle vermutlich ebenso genau vorher wie die Formel der multiplen Regression, die bei der ursprünglichen Stichprobe optimal war. Neuere Forschungen gehen noch weiter: Formeln, die alle Prädiktoren gleich gewichten, sind oftmals überlegen, weil sie nicht von den Zufällen der Stichprobennahme beeinflusst sind.[9]

Der überraschende Erfolg gleich gewichteter Formeln hat eine bedeutende praktische Konsequenz: Es ist möglich, ohne vorherige statistische Datenerhebung nützliche Algorithmen zu entwickeln. Einfache gleich gewichtete Formeln, die auf vorhandenen statistischen Daten oder gesundem Menschenverstand basieren, sind oftmals ausgezeichnete Prädiktoren bedeutsamer Ergebnisse. In einem denkwürdigen Beispiel zeigte Dawes, dass sich die Stabilität einer Ehe gut mit einer Formel vorhersagen lässt:

## Häufigkeit des Geschlechtsverkehrs minus Streithäufigkeit

Sie würden sich nicht wünschen, dass dabei eine negative Zahl herauskommt.

Aus diesen Forschungen lässt sich die wichtige Schlussfolgerung ziehen, dass ein flüchtig konzipierter Algorithmus oftmals gut genug ist, um mit einer optimal gewichteten Formel mitzuhalten – und zweifellos gut genug, um Urteile von Experten zu übertreffen. Diese Logik lässt sich in vielen Bereichen anwenden, angefangen von der Auswahl von Aktien durch Fondsmanager bis hin zur Auswahl von Therapien durch Ärzte oder Patienten.

Eine klassische Anwendung dieser Methode ist ein einfacher Algorithmus, der Hunderttausenden von Kindern das Leben rettete. Geburtshelfer wussten von jeher, dass ein Neugeborenes, das innerhalb weniger Minuten nach der Geburt nicht normal atmet, ein hohes Risiko hat, einen bleibenden Hirnschaden zu erleiden oder zu sterben. Bis die Anästhesistin Virginia Apgar im Jahr 1953 ein neues Verfahren einführte, verließen sich Ärzte und Hebammen auf ihr klinisches Urteil, um festzustellen, ob ein Neugeborenes in akuter Lebensgefahr schwebte. Verschiedene Praktiker konzentrierten sich auf unterschiedliche Symptome. Einige achteten auf Atemprobleme, während andere überwachten, wie schnell das Neugeborene zu schreien anfing. Ohne ein standardisiertes Verfahren wurden Gefahrenzeichen oft übersehen, und viele Neugeborene starben.

Eines Tages beim Frühstück fragte ein Assistenzarzt Dr. Apgar, wie sie den Gesundheitszustand eines Neugeborenen systematisch beurteilen würde.[10] »Das ist leicht«, antwortete sie. »Man würde folgendermaßen vorgehen.« Apgar schrieb fünf Variablen (Herzschlag, Atmung, Reflex, Muskeltonus und Hautfarbe) und drei Punktwerte auf (0, 1 oder 2, je nach der Robustheit jedes Zeichens). Als Apgar erkannte, dass sie womöglich eine bahnbrechende Entdeckung gemacht hatte, die in jedem Kreißsaal angewendet werden konnte, begann sie, Neugeborene eine Minute nach der Geburt nach dieser Regel zu beurteilen. Ein Neugeborenes mit einer Gesamtpunktzahl von mindestens acht war rosig, strampelte, schrie,

grimassierte und hatte einen Herzschlag von mindestens hundert Schlägen pro Minute – war also in einer guten Verfassung. Ein Neugeborenes mit einer Punktzahl von höchstens vier war wahrscheinlich bläulich, schlaff, passiv und hatte einen langsamen oder schwachen Puls – es musste also sofort intensivmedizinisch behandelt werden. Mit dem Apgar-Score verfügte das medizinische Team in Kreißsälen über konsistente Standards zur Beurteilung des Gesundheitszustands von Neugeborenen Die Formel gilt als ein bedeutender Beitrag zur Senkung der Säuglingssterblichkeit. Der Apgar-Test wird noch heute tagtäglich in jedem Kreißsaal angewandt. Atul Gawandes jüngstes Werk *A Checklist Manifesto* liefert viele weitere Beispiele für den Nutzen von Checklisten und einfachen Regeln.[11]

## Die Feindseligkeit gegen Algorithmen

Von Anfang an reagierten klinische Psychologen mit offener Ablehnung und Ungläubigkeit auf Meehls Ideen. Was ihre Fähigkeit zur Erstellung langfristiger Prognosen betraf, erlagen sie eindeutig einer Kompetenzillusion. Bei genauerem Nachdenken ist leicht zu erkennen, wie es zu dieser Illusion kam, und man kann die Ablehnung von Meehls Forschungen durch die Kliniker leicht nachvollziehen.

Die statistischen Belege für die Unterlegenheit der Kliniker widersprechen ihren Alltagserfahrungen bezüglich der Qualität ihrer Urteile. Psychologen, die mit Patienten arbeiten, haben während jeder therapeutischen Sitzung viele Intuitionen, wobei sie antizipieren, wie der Patient auf eine Intervention reagieren wird, und abschätzen, was als Nächstes passieren wird. Viele dieser Intuitionen bestätigen sich, was ihren klinischen Sachverstand scheinbar belegt.

Das Problem ist, dass die richtigen Urteile kurzfristige Vorhersagen im Rahmen des therapeutischen Interviews umfassen – eine Fähigkeit, in der Therapeuten jahrelange Übung haben mögen. Die Aufgaben, in denen sie versagen, erfordern typischerweise langfristige Vorhersagen über die weitere Entwicklung des Patienten. Diese sind viel schwieriger, selbst die besten Formeln liefern nur recht bescheidene Resultate, und es gibt auch Aufgaben, die die Kliniker nie richtig lernen konnten – sie müssten Jahre auf Rückmeldung warten, statt wie sonst während der klinischen Sitzung sofortiges Feedback zu bekommen. Doch die Trennlinie zwischen dem, was Kliniker gut können, und dem, was sie überhaupt nicht können, ist nicht offensichtlich, und erst recht nicht für sie selbst. Sie wissen, dass sie fachkundig sind, aber sie kennen nicht unbedingt die Grenzen ihrer Fachkunde. Da ist es nicht weiter verwunderlich, dass die Vorstellung, eine mechanische Kombination einiger weniger Variablen könne die subtile Komplexität der menschlichen Urteilskraft übertreffen, erfahrenen Klinikern offenkundig absurd erscheint.

Die Debatte über die Vorzüge klinischer und statistischer Vorhersagen hat von jeher eine moralische Dimension gehabt. Die statistische Methode, schrieb Meehl, werde von erfahrenen Klinikern als »mechanisch, atomistisch, additiv, schablonenhaft, künstlich, unwirklich, willkürlich,

unvollständig, tot, pedantisch, zersplittert, trivial, gezwungen, statisch, oberflächlich, starr, steril, akademisch, pseudowissenschaftlich und blind« kritisiert. Die klinische Methode andererseits wurde von ihren Verfechtern als »dynamisch, global, bedeutungsvoll, ganzheitlich, hintergründig, teilnehmend, integrativ, strukturiert, organisiert, reich, tief, echt, gefühlvoll, differenziert, real, lebendig, konkret, natürlich, lebensecht und verständnisvoll« gelobt.

Dies ist eine Einstellung, die wir alle verstehen. Wenn ein Mensch mit einer Maschine konkurriert – ob es der legendäre John Henry war, der beim Stollenbau mit seinem mechanischen Hammer gegen einen dampfbetriebenen Hammer antrat, oder das Schachgenie Garri Kasparow, der gegen den Computer Deep Blue antrat –, gelten unsere Sympathien unserem Mitmenschen. Die Abneigung gegen Entscheidungen von Algorithmen, die sich auf Menschen auswirken, wurzelt in der starken Präferenz vieler Menschen für das Natürliche im Gegensatz zum Künstlichen. Gefragt, ob sie lieber einen Bio-Apfel oder einen Apfel aus herkömmlichem Anbau essen würden, ziehen die meisten Menschen den »natürlichen« Apfel vor. Selbst wenn man sie darüber informiert hat, dass die beiden Äpfel genau gleich schmecken, den gleichen Nährwert haben und gleich gesundheitsfördernd sind, zieht eine Mehrheit weiterhin den Bio-Apfel vor.[12] Selbst Bierproduzenten haben festgestellt, dass sie ihren Absatz steigern können, wenn sie »naturbelassen« oder »ohne Zusatz von Konservierungsstoffen« auf das Etikett schreiben.

Der tiefe Widerstand gegen die Entzauberung des Expertentums zeigt sich in der Reaktion der europäischen Weinliebhaber auf Ashenfelters Formel für die Vorhersage des Preises von Bordeauxweinen. Ashenfelters Formel war ein Geschenk des Himmels: Daher hätte man erwarten sollen, dass ihm Weinliebhaber überall dankbar dafür wären, nachweislich ihre Fähigkeit verbessert zu haben, jene Weine zu identifizieren, die später einmal von vorzüglicher Qualität wären. Weit gefehlt. Die Reaktion in französischen Weinkreisen, schrieb die *New York Times*, »bewegte sich zwischen Empörung und Hysterie«. Ashenfelter berichtet, ein Weinkenner habe seine Ergebnisse als »haarsträubend und absurd« bezeichnet. Ein anderer spottete: »Es ist so, als würde man Kinofilme beurteilen, ohne sie gesehen zu haben.«

Das Vorurteil gegen Algorithmen verstärkt sich, wenn die Entscheidungen folgenreich sind. Meehl bemerkte: »Ich weiß nicht, wie ich

das Entsetzen lindern kann, das einige Kliniker überkommt, wenn sie sich vorstellen, dass einem behandelbaren Fall nur deshalb die Behandlung vorenthalten wird, weil ihn eine ›blinde, mechanische‹ Gleichung falsch klassifiziert.« Dagegen haben Meehl und andere Verfechter des Einsatzes von Algorithmen sehr entschieden behauptet, es sei unethisch, sich bei wichtigen Entscheidungen auf intuitive Urteile zu verlassen, wenn ein Algorithmus verfügbar ist, der weniger Fehler macht. Ihre rationale Argumentation ist schlüssig, aber sie widerspricht einer hartnäckigen psychologischen Tatsache: Für die meisten Menschen ist die Ursache eines Fehlers von Bedeutung. Die Geschichte eines Kindes, das gestorben ist, weil ein Algorithmus einen Fehler gemacht hat, ist ergreifender als die Geschichte derselben Tragödie, die auf einen menschlichen Fehler zurückzuführen ist, und die unterschiedliche emotionale Intensität wird bereitwillig in eine moralische Präferenz übersetzt.

Glücklicherweise wird die Ablehnung von Algorithmen vermutlich in dem Maße zurückgehen, wie sie im Alltagsleben eine immer größere Rolle spielen werden. Auf der Suche nach Büchern oder Musik, die wir mögen, schätzen wir die von Software generierten Empfehlungen. Wir nehmen es als selbstverständlich hin, dass menschliche Urteilskraft keinen direkten Einfluss mehr auf Entscheidungen über Kreditlinien ausübt. Wir sind in zunehmendem Maße Richtlinien ausgesetzt, die die Form einfacher Algorithmen haben, wie etwa das Verhältnis von »gutem« und »schlechtem« Cholesterin, nach dem wir streben sollten. Die Bevölkerung ist sich mittlerweile durchaus der Tatsache bewusst, dass bei einigen kritischen Entscheidungen in der Welt des Sports Formeln leistungsfähiger sind als Menschen: Wie viel sollte ein Profiteam für bestimmte Nachwuchsspieler zahlen, oder wann sollte man beim American Football beim vierten Versuch einen sogenannten Punt-Spielzug versuchen? Die stetig länger werdende Liste von Aufgaben, mit denen Algorithmen betraut werden, sollte langfristig das Unbehagen vermindern, das die meisten Menschen empfinden, wenn sie zum ersten Mal dem Muster der Ergebnisse begegnen, das Meehl in seinem verstörenden kleinen Buch beschrieben hat.

## Von Paul Meehl lernen

Als 21-jähriger Leutnant in der israelischen Armee sollte ich 1955 ein Interviewsystem für die gesamte Armee entwickeln. Falls Sie sich fragen sollten, wieso einem so jungen Offizier eine so verantwortungsvolle Aufgabe übertragen wurde, sollten Sie sich daran erinnern, dass der Staat Israel selbst damals erst sieben Jahre alt war; all seine Institutionen befanden sich im Aufbau, und jemand musste sie aufbauen. So seltsam es sich heute anhören mag, hat mich mein Bachelorabschluss in Psychologie vermutlich zum am besten ausgebildeten Psychologen in der Armee gemacht. Mein direkter Vorgesetzter, ein brillanter Forscher, war Chemiker.

Als man mir diese Aufgabe übertrug, gab es bereits ein bestimmtes Befragungsprozedere. Jeder Soldat wurde bei seiner Einberufung einer Reihe psychometrischer Tests unterzogen, und jeder Rekrut, der für Kampfeinsätze in Betracht gezogen wurde, wurde interviewt, um ein Persönlichkeitsprofil zu erstellen. Ziel war es, dem Rekruten einen Punktwert für seine allgemeine Kampftauglichkeit zuzuschreiben und die optimal zu seiner Persönlichkeit passende Verwendung zu finden: Infanterie, Artillerie, Panzertruppen und so weiter. Die Interviewer waren ihrerseits junge Wehrpflichtige, die wegen ihrer hohen Intelligenz und ihres Interesses am Umgang mit Menschen für diese Aufgabe ausgewählt worden waren. Die meisten waren Frauen, die damals noch vom Dienst an der Waffe ausgenommen waren. Sie wurden ein paar Wochen lang darin geschult, wie man ein 15- bis 20-minütiges Interview führt; sie sollten dabei eine breite Palette von Themen ansprechen und sich einen allgemeinen Eindruck darüber verschaffen, wie gut der Rekrut in der Armee zurechtkommen würde.

Leider hatten anschließende Evaluierungen bereits darauf hingedeutet, dass diese Befragungsmethode für die Vorhersage des zukünftigen Erfolgs der Rekruten fast nutzlos war. Ich sollte einen Interviewfragebogen erstellen, der nützlicher war, aber nicht mehr Zeit beanspruchte. Ich sollte den neuen Fragebogen auch erproben und seine Vorhersagegenauigkeit beurteilen. Aus der Sicht eines Fachmanns war ich für diese Aufgabe nicht besser qualifiziert, als ich es gewesen wäre, um eine Brücke über den Amazonas zu bauen.

Zum Glück hatte ich Paul Meehls »kleines Buch« gelesen, das ein Jahr zuvor veröffentlicht worden war. Seine Behauptung, einfache statistische Regeln seien intuitiven »klinischen« Urteilen überlegen, hatte mich überzeugt. Ich gelangte zu dem Schluss, dass der damals gebräuchliche Fragebogen zumindest zum Teil deshalb ein Fehlschlag war, weil er den Interviewern erlaubte, das zu tun, was sie am interessantesten fanden, nämlich etwas über das innere Seelenleben des Befragten zu erfahren. Vielmehr sollten wir die begrenzte Zeit, die uns zur Verfügung stand, dazu benutzen, möglichst viele spezifische Informationen über das Leben des Befragten in seinem normalen Umfeld in Erfahrung zu bringen. Darüber hinaus lernte ich von Meehl, dass die endgültige Entscheidung nicht länger von den allgemeinen Beurteilungen des Rekruten durch den Interviewer bestimmt werden sollte. Meehls Buch suggerierte, dass man solchen Beurteilungen nicht vertrauen konnte und dass statistische Zusammenfassungen von getrennt bewerteten Attributen eine höhere prognostische Gültigkeit besitzen.

Ich entschied mich für ein Verfahren, bei dem die Interviewer mehrere relevante Persönlichkeitszüge beurteilen und jedem für sich einen Punktwert zuschreiben sollten. Die Gesamtpunktzahl für die Kampfeinsatztauglichkeit sollte nach einer Standardformel berechnet werden, ohne dass die Interviewer weiter Einfluss darauf nehmen sollten. Ich erstellte eine Liste von sechs Merkmalen, die für die Leistung in einer Kampfeinheit relevant zu sein schienen, darunter »Verantwortungsbewusstsein«, »Umgänglichkeit« und »männlicher Stolz«. Anschließend entwarf ich für jeden Persönlichkeitszug eine Reihe von Sachfragen, die sich auf das Leben des Rekruten vor seiner Einberufung bezogen, wie etwa die Zahl der bisherigen Anstellungen, wie regelmäßig und pünktlich er seiner Arbeit oder seinem Studium nachgegangen ist, die Häufigkeit seiner Kontakte mit Freunden, sein Interesse für und seine Teilnahme an sportlichen Aktivitäten und so weiter. Es ging darum, so objektiv wie möglich zu beurteilen, wie gut der jeweilige Rekrut in jeder Dimension abgeschnitten hatte.

Durch die Konzentration auf standardisierte, sachbezogene Fragen wollte ich den Halo-Effekt reduzieren, bei dem günstige erste Eindrücke spätere Urteile beeinflussen. Als weitere Schutzvorkehrung gegen Halo-Effekte wies ich die Interviewer an, die sechs Persönlichkeitsmerkmale in einer festen Reihenfolge durchzugehen und jedes Merkmal auf einer Fünf-

Punkte-Skala zu beurteilen, ehe sie zum nächsten weitergingen. Und das war es. Ich sagte den Interviewern, sie sollten sich nicht mit der Frage befassen, wie gut sich der Rekrut in die militärische Organisation einfügen würde. Sie sollten lediglich sachdienliche Tatsachen über seine Vergangenheit eruieren und diese Informationen dazu benutzen, dem Rekruten in jeder Persönlichkeitsdimension einen bestimmten Punktwert zuzuschreiben. »Ihre Aufgabe ist es, zuverlässige Messwerte bereitzustellen«, sagte ich ihnen. »Überlassen Sie die Frage der prognostischen Gültigkeit mir«, womit ich die Formel meinte, die ich konzipieren wollte, um ihre jeweiligen Ratings zu kombinieren.

Die Interviewer hätten beinahe gemeutert. Diesen intelligenten jungen Leuten missfiel es, dass ihnen jemand, der kaum älter war als sie selbst, befahl, ihre Intuition auszuschalten und sich ganz und gar auf langweilige Sachfragen zu konzentrieren. Eine von ihnen beklagte: »Sie machen uns zu Robotern!« Und so machte ich ihnen ein Zugeständnis. »Führen Sie das Interview genauso durch, wie ich Sie angewiesen habe«, sagte ich ihnen, »und wenn Sie das getan haben, erfüllen Sie sich Ihren Wunsch: Schließen Sie die Augen, stellen Sie sich den Rekruten als Soldaten vor und schreiben Sie ihm auf einer Skala von eins bis fünf einen Punktwert zu.«

Mehrere Hundert Interviews wurden nach dieser neuen Methode geführt, und einige Monate später sammelten wir die Leistungsbewertungen der Soldaten von den befehlshabenden Offizieren der Einheiten, denen sie zugeteilt worden waren. Die Ergebnisse machten uns glücklich. Wie Meehls Buch nahegelegt hatte, stellte das neue Interviewverfahren eine erhebliche Verbesserung gegenüber dem alten dar. Die Summe unserer sechs Ratings sagte die Leistung der Soldaten viel genauer vorher als die globalen Beurteilungen der vorhergehenden Befragungsmethode, wenn auch keineswegs vollkommen. Wir hatten uns von »völlig nutzlos« auf »mittelmäßig nützlich« verbessert.

Die große Überraschung für mich bestand darin, dass das intuitive Urteil, zu dem die Interviewer in der »Schließ-die-Augen-Übung« gelangten, ebenfalls eine recht gute Prognose lieferte, die tatsächlich genauso gut war wie die Summe der sechs spezifischen Ratings. Aus diesem Befund lernte ich eine Lektion, die ich nie mehr vergessen habe: Intuition liefert einen Mehrwert, selbst in dem zu Recht verspotteten Auswahlinterview, aber erst nach einer disziplinierten Erhebung objektiver Informationen und der disziplinierten Bewertung gesonderter Merkmale. Ich stellte eine Formel

auf, die der Beurteilung »mit geschlossenen Augen« das gleiche Gewicht gab wie der Summe der sechs Merkmalsratings. Aus dieser Episode lernte ich auch noch die allgemeine Lektion, nicht einfach dem intuitiven Urteil – dem eigenen oder dem anderer – zu vertrauen, aber es auch nicht als belanglos abzutun.

Etwa 45 Jahre später, nachdem ich den Nobelpreis für Wirtschaftswissenschaften erhalten hatte, war ich kurzzeitig eine kleine Berühmtheit in Israel. Bei einem meiner Besuche hatte jemand die Idee, mich durch meinen ehemaligen Armeestützpunkt zu führen, in dem noch immer die Einheit untergebracht war, die für das Auswahlverfahren der Rekruten zuständig war. Ich wurde der befehlshabenden Offizierin der Psychologischen Einheit vorgestellt, und sie schilderte ihr gegenwärtiges Interviewverfahren, das sich nicht stark von dem System unterschied, das ich entwickelt hatte. Wie sich herausstellte, deuteten zahlreiche Forschungsergebnisse darauf hin, dass die Interviews noch immer gute Dienste leisteten. Als die Offizierin ans Ende ihrer Beschreibung der Interviewführung gelangte, sagte sie noch: »Und dann sagen wir ihnen: ›Schließen Sie die Augen.‹«

## Probieren Sie es selbst aus

Die Botschaft dieses Kapitels lässt sich leicht auch auf andere Aufgaben als Personalentscheidungen einer Armee anwenden. Die Umsetzung von Interviewverfahren im Geist von Meehl und Dawes verlangt vergleichsweise wenig Anstrengung, aber sehr viel Disziplin. Angenommen, Sie müssen einen Vertreter für Ihre Firma einstellen. Wenn Sie wirklich die bestmögliche Person für den Job bekommen wollen, sollten Sie folgendermaßen vorgehen. Wählen Sie, erstens, einige Persönlichkeitsmerkmale aus, die Voraussetzungen für den Erfolg in dieser Position sind (fachliche Befähigung, einnehmende Persönlichkeit, Zuverlässigkeit und so weiter). Übertreiben Sie nicht – sechs Dimensionen sind eine gute Zahl. Die Merkmale, die Sie auswählen, sollten so unabhängig wie möglich voneinander sein, und Sie sollten das Gefühl haben, dass Sie sie zuverlässig bewerten können, indem sie ein paar sachbezogene Fragen stellen. Erstellen Sie als Nächstes eine Liste dieser Fragen für jedes Merkmal, und überlegen Sie, wie Sie es bewerten wollen, zum Beispiel auf einer Skala von eins bis fünf. Sie sollten eine Vorstellung davon haben, was Sie unter »sehr schwach« oder »sehr stark« verstehen.

Diese Vorbereitungen sollten Sie etwa eine halbe Stunde kosten, eine kleine Investition, die in Bezug auf die Qualität der Mitarbeiter, die Sie einstellen, einen großen Unterschied machen kann. Um Halo-Effekte zu vermeiden, sollten Sie die Informationen über die einzelnen Merkmale nacheinander zusammentragen und jedes einzelne bewerten, ehe Sie zum nächsten weitergehen. Überspringen Sie keine Merkmale. Um jeden Kandidaten zu bewerten, zählen Sie die sechs Punktwerte zusammen. Weil Sie für die endgültige Entscheidung verantwortlich sind, sollten Sie sich die Kandidaten nicht »mit geschlossenen Augen« vorstellen. Sie sollten fest entschlossen den Kandidaten einstellen, dessen Gesamtpunktzahl am höchsten ist, selbst wenn Sie einen anderen lieber mögen – widerstehen Sie dem Wunsch, »gebrochene Beine« zu erfinden, um das Ranking zu ändern. Zahlreiche Forschungsergebnisse sind vielversprechend: Die Wahrscheinlichkeit, dass Sie mit diesem Verfahren den besten Kandidaten auswählen, ist viel höher, als wenn Sie das tun, was Menschen normalerweise in solchen Situationen tun, nämlich unvorbereitet in das

Interview zu gehen und nach einem intuitiven Pauschalurteil wie »Ich sah ihm in die Augen und mochte, was ich sah« eine Auswahl zu treffen.

## Zum Thema »Intuitionen und Formeln«

»Immer wenn wir menschliche Urteile durch eine Formel ersetzen können, sollten wir dies zumindest in Betracht ziehen.«

»Er hält seine Urteile für vielschichtig und scharfsinnig, aber eine einfache Kombination von Scores dürfte ein verlässlicheres Ergebnis bringen.«

»Wir sollten vorab entscheiden, wie wir die Daten gewichten wollen, die wir über die vergangenen Leistungen der Kandidaten besitzen. Andernfalls geben wir unserem Eindruck aus den Interviews zu viel Gewicht.«

# *22. Die Intuition von Experten: Wann können wir ihr vertrauen?*

Kontroversen unter Fachleuten bringen die schlechtesten Seiten der akademischen Welt zum Vorschein. Wissenschaftsmagazine veröffentlichen gelegentlich einen Schlagabtausch zwischen Forschern, der oftmals damit beginnt, dass jemand die Forschungen eines anderen kritisiert, gefolgt von einer Erwiderung und einer Gegenerwiderung. Ich habe diese Wortgefechte immer für Zeitvergeudung gehalten. Insbesondere wenn die ursprüngliche Kritik in scharfe Worte gefasst ist, sind die Erwiderung und die Gegenerwiderung oftmals Übungen in dem, was ich »Sarkasmus für Anfänger und Fortgeschrittene« genannt habe. Die Antworten machen gegenüber einer harschen Kritik nur selten Zugeständnisse, und es ist wohl noch nie vorgekommen, dass eine Gegenerwiderung einräumt, dass die ursprüngliche Kritik verfehlt oder in irgendeiner Weise irrig gewesen sei. Einige wenige Male habe ich auf Kritik reagiert, die ich für grob irreführend hielt, weil es als Eingeständnis eines Fehlers interpretiert werden kann, wenn man nicht reagiert, aber ich habe diese feindseligen Dispute nie als lehrreich erlebt. Auf der Suche nach einem anderen Weg, eine Meinungsverschiedenheit auszutragen, habe ich einige Male an *adversarial collaborations* (»gegnerischen Kooperationen«) mitgewirkt, bei denen sich Forscher mit gegensätzlichen wissenschaftlichen Standpunkten bereit erklären, gemeinsam einen Aufsatz über ihre Differenzen zu schreiben oder auch ein gemeinsames Forschungsprojekt durchzuführen. In besonders angespannten Situationen wird das Forschungsprojekt von einem Vermittler moderiert.[1]

Meine befriedigendste und fruchtbarste »gegnerische Kooperation« war mit Gary Klein, dem geistigen Kopf einer Vereinigung von Wissenschaftlern und Praktikern, die meine Arbeit ablehnen. Sie nennen sich selbst Erforscher der »Naturalistischen Entscheidungsfindung« (*Naturalistic Decision Making*, NDM), und sie arbeiten hauptsächlich in Organisationen, in denen sie oftmals die Arbeitsweise von Experten erforschen. Die NDM-Anhänger lehnen den Fokus auf kognitive

Verzerrungen im Rahmen des Heuristik- und Verzerrungsansatzes entschieden ab. Sie kritisieren an diesem Modell seine einseitige Ausrichtung auf Urteilsfehler und auf künstliche Experimente sowie seine vermeintliche Vernachlässigung der Erforschung realer Menschen, die wirklich belangvolle Dinge tun. Sie beurteilen den Nutzen starrer Algorithmen als Ersatz für menschliche Urteilskraft äußerst skeptisch, und Paul Meehl gehört nicht zu ihren Helden. Gary Klein hat diese Position viele Jahre lang mit großer Beredsamkeit vertreten.[2]

Dies ist kaum die Basis für eine innige Freundschaft, aber es hatte mehr damit auf sich. Ich war nie der Ansicht, Intuition führe immer in die Irre. Ich war ein Fan von Kleins Studien über die Expertise (das Expertenwissen) von Feuerwehrleuten, seit ich zum ersten Mal den Entwurf eines Aufsatzes sah, den er in den 1970er-Jahren schrieb. Und ich war beeindruckt von seinem Buch *Natürliche Entscheidungsprozesse. Über die »Quellen der Macht«, die unsere Entscheidungen lenken*, in dem er ergründet, wie erfahrene Fachleute intuitive Fähigkeiten entwickeln. Ich lud ihn ein, sich an einem Forschungsvorhaben zu beteiligen, bei dem es darum gehen sollte, die Grenze zwischen den wunderbaren Leistungen der Intuition und ihren Fehlern genauer abzustecken. Die Idee faszinierte ihn, und wir nahmen das Projekt in Angriff – ohne die Gewissheit, dass es funktionieren würde. Wir nahmen uns vor, eine ganz konkrete Frage zu beantworten: Wann kann man einem erfahrenen Fachmann vertrauen, der behauptet, eine sichere Intuition zu haben? Es lag auf der Hand, dass Klein eher bereit wäre, ihm zu vertrauen, und dass ich skeptischer wäre. Aber könnten wir uns auf Prinzipien für die Beantwortung der allgemeinen Frage verständigen?

Im Verlauf von sieben oder acht Jahren hatten wir viele Diskussionen, legten viele Differenzen bei, standen mehr als einmal kurz davor, alles hinzuschmeißen, schrieben viele Entwürfe, wurden Freunde und veröffentlichten schließlich einen gemeinsamen Artikel mit einem Titel, der alles sagt: »Bedingungen für intuitive Expertise: fehlende Uneinigkeit« (»Conditions for Intuitive Expertise: A Failure to Disagree«) Tatsächlich waren wir auf keine grundlegende Frage gestoßen, in der wir unterschiedlicher Meinung waren – aber wir waren uns auch nicht wirklich einig.

## Glanzleistungen und Mängel

Malcolm Gladwells Bestseller *Blink!* erschien, als Klein und ich an dem Projekt arbeiteten, und es war beruhigend, herauszufinden, dass wir uns über das Werk einig waren. Gladwells Buch beginnt mit der denkwürdigen Geschichte von Kunstexperten, denen ein Objekt präsentiert wurde, das als ein prächtiger Kouros, die Statue eines schreitenden Jünglings, beschrieben wurde.[3] Mehrere der Experten hatten aus dem Bauch heraus das starke Gefühl, dass die Statue gefälscht war, aber sie konnten nicht in Worte fassen, woran sich dieses Unbehagen entzündete. Jeder, der dieses Buch las – Millionen taten es –, erinnert sich an diese Geschichte als einen Triumph der Intuition. Sie scheint den Eindruck zu vermitteln, dass eine systematische Suche nach dem Hinweisreiz, der die Experten leitete, umsonst gewesen wäre, aber Klein und ich wiesen diese Schlussfolgerung zurück. Aus unserer Sicht war eine solche Untersuchung notwendig, und wenn sie sachgerecht durchgeführt worden wäre (worauf sich Klein versteht), wäre sie wahrscheinlich erfolgreich gewesen.

Obgleich viele Leser des Kouros-Beispiels bestimmt zu einer fast magischen Sicht der Intuition von Experten gelangten, ist Gladwell selbst nicht dieser Meinung. In einem späteren Kapitel beschreibt er ein massives Versagen der Intuition: Die Amerikaner wählten Präsident Harding, dessen einzige Qualifikation für das Amt darin bestand, dass er rein äußerlich für die Rolle wie geschaffen schien. Mit seiner stattlichen Figur und seinem breiten Kinn verkörperte er das Ideal einer starken und entschlossenen Führungspersönlichkeit. Die Menschen wählten jemanden, der stark und entschlossen aussah, ohne dass sie irgendwelche sonstigen Anhaltspunkte dafür hatten, dass er dies tatsächlich war. Eine intuitive Vorhersage über die Tüchtigkeit Hardings als Präsident war das Resultat der Ersetzung einer Frage durch eine andere. Ein Leser dieses Buches sollte erwarten, dass an einer solchen Intuition mit großer innerer Überzeugung festgehalten wird.

## Intuition als Wiedererkennen

Die frühen Erfahrungen, die Kleins Sichtweise der Intuition prägten, unterschieden sich grundlegend von meinen eigenen. Mein Denken wurde maßgeblich durch die Erfahrung beeinflusst, dass ich die Illusion der Prognosegültigkeit bei mir beobachtete und Paul Meehls Demonstrationen der Unterlegenheit klinischer Vorhersagen las. Kleins Ansichten dagegen wurden von seinen frühen Studien über Einsatzleiter der Feuerwehr beeinflusst. Er folgte ihnen, wenn sie Brände bekämpften, und befragte den Einsatzleiter später darüber, welche Überlegungen seinen Entscheidungen zugrunde lagen. Klein schrieb in unserem gemeinsamen Artikel:

> Ich untersuchte mit meinen Mitarbeitern, wie die Einsatzleiter, ohne Optionen zu vergleichen, gute Entscheidungen treffen konnten. Die Ausgangshypothese lautete, die Einsatzleiter würden sich bei ihrer Analyse auf nur ein paar Optionen beschränken, aber diese Hypothese erwies sich als falsch. Tatsächlich generierten die Einsatzleiter für gewöhnlich nur eine Option, und das war alles, was sie brauchten. Sie konnten sich auf das Repertoire von Verhaltensmustern stützen, das sie in über einem Jahrzehnt realer und virtueller Erfahrungen zusammengetragen hatten, um eine plausible Option zu ermitteln, die sie als Erstes in Betracht zogen. Sie bewerteten diese Option, indem sie sie mental simulierten, um herauszufinden, ob sie in der Situation, in der sie sich befanden, Erfolg versprechend wäre … Wenn die Vorgehensweise, die sie in Betracht zogen, geeignet zu sein schien, setzten sie sie um. Wenn sie Unzulänglichkeiten aufwies, modifizierten sie sie. Wenn sie sich nicht leicht modifizieren ließ, wandten sie sich der zweitplausibelsten Option zu und spielten sie in der gleichen Weise durch, bis eine geeignete Vorgehensweise gefunden wurde.

Klein arbeitete diese Beschreibung zu einer Theorie der Entscheidungsfindung aus, die er »RPD-Modell« *(recognition-primed decision model,* »Modell der wiedererkennungsgesteuerten Entscheidung«) nannte, das für Feuerwehrleute gilt, aber auch Expertise auf anderen Feldern einschließlich Schach beschreibt. Der Entscheidungsprozess bezieht sowohl System 1 als auch System 2 ein. In der ersten Phase wird durch eine automatische Funktion des assoziativen Gedächtnisses – System 1 – ein vorläufiger Plan ins Bewusstsein gehoben. Die nächste Phase ist ein zielgerichteter, willentlicher Prozess, in dem der Plan mental simuliert wird, um seine Effektivität zu überprüfen – eine Operation von System 2. Das Modell der intuitiven Entscheidungsfindung als Musterwiedererkennung entwickelt Ideen weiter, die Herbert Simon schon vor einiger Zeit vorlegte – der vielleicht einzige Wissenschaftler, der von allen miteinander konkurrierenden Gruppen auf dem Feld der Entscheidungsforschung als Held anerkannt und bewundert wird.[4] Ich zitierte Herbert Simons Definition der Intuition in der Einleitung, aber ich möchte sie wiederholen, weil wir sie jetzt besser verstehen: »Die Situation liefert einen Hinweisreiz *(cue);* dieser Hinweisreiz gibt dem Experten Zugang zu Informationen, die im Gedächtnis gespeichert sind, und diese Informationen geben ihm die Antwort. Intuition ist nicht mehr und nicht weniger als Wiedererkennen.«[5]

Diese prägnante Aussage reduziert die scheinbare Magie der Intuition auf die Alltagserfahrung des Gedächtnisses. Wir staunen über die Geschichte des Feuerwehrmanns, der plötzlich den Drang verspürt, aus einem brennenden Haus zu fliehen, kurz bevor es einstürzt, weil der Feuerwehrmann die Gefahr intuitiv erkennt, »ohne zu wissen, woher er es weiß«.[6] Allerdings wissen wir auch nicht, wieso wir sofort wissen, dass eine Person, die wir beim Betreten eines Raumes erblicken, unser Freund Peter ist. Die Moral von Simons Bemerkung lautet, dass das Rätsel des »Wissens, ohne zu wissen« kein charakteristisches Merkmal der Intuition ist; es ist die Norm des mentalen Lebens.

## Erwerb von Fertigkeiten

Wie werden die Informationen, die die Grundlage von Intuitionen bilden, »im Gedächtnis gespeichert«? Bestimmte Typen von Intuitionen werden sehr schnell erworben. Von unseren Ahnen haben wir die Fähigkeit geerbt, sehr schnell zu lernen, wann wir uns fürchten müssen. Tatsächlich ist oftmals schon eine einzige Erfahrung ausreichend, um eine langfristige Abneigung und Furcht zu erzeugen. Viele von uns haben eine tief sitzende Erinnerung an ein einziges fragwürdiges Gericht, das uns noch immer instinktiv davon abhält, in ein Restaurant zurückzukehren. Wir alle erleben eine starke Anspannung, wenn wir uns einer Stelle nähern, an der sich etwas Unangenehmes ereignet hat, auch wenn es keinen Grund zur Annahme gibt, dass es sich wiederholen wird. Ein solcher Ort ist für mich die Zufahrt zum Flughafen von San Francisco, wo mir vor Jahren ein aggressiver Fahrer von der Fernstraße nachfuhr, das Fenster runterkurbelte und mich wüst beschimpfte. Ich habe nie erfahren, was seinen Hass auslöste, aber jedes Mal, wenn ich auf dem Weg zum Flughafen an dieser Stelle vorbeikomme, erinnere ich mich an seine Stimme.

Die Erinnerung an den Vorfall am Flughafen ist mir bewusst, und sie erklärt vollständig das Gefühl, das damit verbunden ist. Doch bei zahlreichen Gelegenheiten mögen wir uns an einem bestimmten Ort unwohl fühlen, oder wenn jemand einen bestimmten Ausdruck verwendet, ohne dass wir eine bewusste Erinnerung an das auslösende Ereignis haben. Im Nachhinein nennen wir dieses mulmige Gefühl Intuition, wenn eine negative Erfahrung darauf folgt. Zwischen dieser Art des emotionalen Lernens und dem, was in den berühmten Konditionierungsexperimenten von Pawlow geschah, in denen Hunde lernten, den Klang einer Glocke als Signal für die bevorstehende Fütterung zu erkennen, besteht ein enger Zusammenhang. Was Pawlows Hunde lernten, lässt sich als antrainierte Hoffnung beschreiben. Ängste werden sogar noch leichter erlernt.

Furcht kann auch – sogar recht leicht – durch Wörter statt durch Erfahrungen gelernt werden. Der Feuerwehrmann mit dem »sechsten Sinn« für Gefahren hatte zweifellos zahlreiche Gelegenheiten, bei denen er über die Arten von Bränden, mit denen er persönlich noch nicht konfrontiert war, diskutieren und nachdenken und im Geist wiederholen konnte, welche

Hinweisreize auftreten können und wie er reagieren sollte. Wie ich aus eigener Erfahrung weiß, nimmt die Anspannung eines jungen Zugführers ohne Kampferfahrung stark zu, wenn er Soldaten durch eine enger werdende Schlucht führt, weil ihm beigebracht wurde, dass ein solches Gelände günstig für einen Hinterhalt ist. Um zu lernen, bedarf es keiner häufigen Wiederholungen.

Das emotionale Lernen mag rasch erfolgen, aber die Aneignung von Expertise erfordert im Allgemeinen viel Zeit. Der Erwerb von Expertise bei komplexen Aufgaben wie hochkarätigem Schach, professionellem Basketball oder Brandbekämpfung ist mühsam und langwierig, weil sie auf einem Gebiet nicht aus einer einzelnen Fähigkeit besteht, sondern eher aus einer großen Gesamtheit von Minikompetenzen. Schach ist ein gutes Beispiel. Ein Profi kann eine komplexe Stellung auf einen Blick durchschauen, aber es dauert Jahre, diese Fähigkeit zu entwickeln. Studien über Schachmeister haben gezeigt, dass mindestens 10 000 Stunden konzentrierter Übung (dazu muss man etwa sechs Jahre lang täglich fünf Stunden Schach spielen) erforderlich sind, das höchste Leistungsniveau zu erreichen.[7] Während dieser Stunden intensiver Konzentration wird ein Schachspieler mit Tausenden von Konfigurationen vertraut, die jeweils aus einer Anordnung zusammenhängender Figuren bestehen, welche sich gegenseitig bedrohen oder verteidigen.

Das Erlernen von Schach auf höchstem Niveau lässt sich mit dem Erlernen der Lesefähigkeit vergleichen. Ein Erstklässler muss sich sehr anstrengen, um einzelne Buchstaben zu erkennen und sie zu Silben und Wörtern zusammenzustellen, dagegen nimmt ein guter erwachsener Leser ganze Sätze wahr. Ein sachkundiger Leser hat auch die Fähigkeit erworben, vertraute Elemente zu einem neuen Muster zusammenzufügen, und er kann ein Wort, das er noch nie zuvor gesehen hat, schnell »wiedererkennen« und richtig aussprechen. Beim Schach spielen wiederkehrende Muster interagierender Figuren die Rolle von Buchstaben, und eine Schachposition ist ein langes Wort oder ein Satz.

Ein versierter Leser, der die erste Strophe von Lewis Carrolls Gedicht »Jabberwocky« zum ersten Mal sieht, kann sie mit perfekter Rhythmisierung und Intonation vorlesen und dabei auch noch Spaß haben:

*’Twas brillig, and the slithy toves*
*Did gyre and gimble in the wabe:*
*All mimsy were the borogoves,*
*And the mome raths outgrabe.*

Verdaustig war’s und glasse Wieben
rotterten gorkicht im Gemank;
Gar elump war der Pluckerwank,
Und die gabben Schweisel frieben.
(Dt. von Christian Enzensberger)

Es ist mühsamer und langwieriger, im Schach Expertise zu erwerben, als lesen zu lernen, weil das »Alphabet« des Schachs viel mehr Buchstaben enthält und weil die »Wörter« aus vielen Buchstaben bestehen. Nach Tausenden von Stunden Übung können Schachmeister dann allerdings eine Spielsituation auf einen Blick erfassen. Die wenigen Züge, die ihnen einfallen, sind fast immer stark und manchmal kreativ. Sie können mit einem »Wort« zurechtkommen, dem sie noch nie begegnet sind, und sie können eine neue Interpretation für ein vertrautes Wort finden.

## Die geeignete Umgebung für Expertise

Klein und ich stellten sehr schnell fest, dass wir uns über die Natur intuitiver Fähigkeiten und ihren Erwerb einig waren. Aber in unserer Schlüsselfrage mussten wir uns noch verständigen: Wann kann man einem selbstbewusst auftretenden Experten, der behauptet, eine Intuition zu haben, vertrauen?

Wir gelangten schließlich zu der Überzeugung, dass unsere Meinungsverschiedenheit zum Teil auf die Tatsache zurückzuführen war, dass wir an verschiedene Experten dachten. Klein hatte viel Zeit mit Einsatzleitern der Feuerwehr, Krankenpflegern und anderen Fachkräften verbracht, die über echte Expertise verfügten. Ich hatte mehr Zeit damit verbracht, über Klinikärzte, Stockpicker und Politikwissenschaftler nachzudenken, die nicht tragfähige langfristige Vorhersagen zu machen versuchten. Da war es nicht weiter verwunderlich, dass seine Standardeinstellung Vertrauen und Respekt war, während die meine Skepsis war. Er war eher bereit, Experten zu glauben, die behaupteten, eine intuitive Lösung gefunden zu haben, weil wahre Experten, wie er mir sagte, die Grenzen ihres Wissens kennen. Ich behauptete, dass es viele Pseudoexperten gebe, die nicht ahnen, dass sie nicht wissen, was sie tun (die Illusion der Prognosegültigkeit), und dass grundsätzlich der Grad der subjektiven Überzeugung für gewöhnlich zu hoch und oftmals ohne informativen Gehalt ist.

Weiter vorn führte ich den Grad des subjektiven Überzeugtseins auf zwei miteinander verbundene Eindrücke zurück: kognitive Leichtigkeit und Kohärenz. Wir sind überzeugt, wenn uns die Geschichte, die wir uns selbst erzählen, mühelos einfällt, widerspruchsfrei daherkommt und kein konkurrierendes Szenario vorhanden ist. Aber Leichtigkeit und Kohärenz sind keine Garantie dafür, dass eine für wahr gehaltene Überzeugung tatsächlich wahr ist. Die Assoziationsmaschine ist so eingestellt, dass sie Zweifel unterdrückt und Ideen und Informationen, die mit der aktuell dominanten Geschichte vereinbar sind, ins Gedächtnis ruft. Ein Intellekt, der der Regel folgt, nur die aktuell verfügbaren Informationen zu berücksichtigen, wird viel zu leicht einen hohen Grad des Überzeugtseins erreichen, indem er alles ausblendet, was er nicht weiß. Es ist daher nicht

überraschend, dass viele von uns allzu sehr von der Richtigkeit unbegründeter Intuitionen überzeugt sind. Klein und ich verständigten uns schließlich auf ein wichtiges Prinzip: Das Vertrauen, das Menschen in ihre Intuitionen haben, ist kein verlässlicher Maßstab für deren Richtigkeit. Anders gesagt, trauen Sie niemandem – auch nicht sich selbst –, der Ihnen sagt, dass Sie seinem Urteil vertrauen sollten.

Wenn man der subjektiven Überzeugung nicht trauen kann, wie können wir dann die wahrscheinliche Gültigkeit einer intuitiven Aussage beurteilen? Wann spiegelt sich in Urteilen wahre Expertise wider? Wann manifestiert sich in ihnen eine »Illusion der Gültigkeit«? Die Antwort darauf lautet, dass zwei grundlegende Voraussetzungen für den Erwerb von Expertise erfüllt sein müssen:

– Eine Umgebung, die hinreichend regelmäßig ist, um vorhersagbar zu sein.
– Eine Gelegenheit, diese Regelmäßigkeiten durch langjährige Übung zu erlernen.

Wenn diese beiden Bedingungen erfüllt sind, sind Intuitionen vermutlich sachgerecht. Schach ist das Extrembeispiel einer regulären Umgebung, aber Bridge und Poker weisen ebenfalls robuste statistische Regelmäßigkeiten auf, die den Erwerb von Sachverstand ermöglichen. Ärzte, Pfleger, Sportler und Feuerwehrleute sind auch mit komplexen, aber grundsätzlich geordneten Situationen konfrontiert. Die zutreffenden Intuitionen, die Gary Klein beschrieben hat, sind darauf zurückzuführen, dass System 1 des Experten gelernt hat, Hinweisreize mit hoher prognostischer Gültigkeit zu nutzen, auch wenn System 2 nicht gelernt hat, sie zu benennen. Stockpicker und Politikwissenschaftler operieren dagegen in einer Umgebung, in der Informationen keinerlei prognostische Aussagekraft besitzen *(zero-validity environment).* In ihrem Versagen spiegelt sich die grundlegende Nichtvorhersagbarkeit der Ereignisse wider, die sie zu prognostizieren versuchen.

Einige Umgebungen sind nicht bloß unregelmäßig, sondern sogar regelrecht verhängnisvoll. Robin Hogarth beschrieb »fatale« Umgebungen, in denen die Erfahrung Experten mit hoher Wahrscheinlichkeit die falschen Lektionen lehrt. Er entlehnt von Lewis Thomas das Beispiel eines Arztes zu Beginn des 20. Jahrhunderts, der oftmals intuitiv zu ahnen glaubte, wann Patienten im Begriff waren, Typhus zu entwickeln. Leider überprüfte er

seine Ahnungen, indem er die Zungen der verschiedenen Patienten abtastete, ohne sich dazwischen die Hände zu waschen. Als ein Patient nach dem anderen erkrankte, entwickelte der Arzt ein Gefühl klinischer Unfehlbarkeit. Seine Vorhersagen waren zutreffend – aber nicht weil er über eine professionelle Intuition verfügte!

Meehls Kliniker waren nicht unfähig, und ihr Versagen war nicht auf einen Mangel an Talent zurückzuführen. Ihre Erfolgsbilanz war deshalb so mager, weil ihnen Aufgaben zugewiesen wurden, die keine einfachen Lösungen hatten. Die Kliniker operierten in einem Umfeld, das nicht ganz so extrem war wie das *zero-validity environment* langfristiger politischer Vorhersagen, aber sie waren mit Situationen konfrontiert, die keine sehr genauen Prognosen erlaubten. Wir wissen dies, weil die besten statistischen Algorithmen, obgleich sie eine höhere Genauigkeit besitzen als menschliche Urteile, nicht sehr genau sind. Tatsächlich haben die Studien Meehls und seiner Nachfolger nie einen Fall nachgewiesen, in dem die Kliniker einen Hinweis mit hoher prognostischer Gültigkeit, den der Algorithmus erfasste, vollständig übersahen. Ein extremes Versagen dieser Art ist unwahrscheinlich, weil menschliches Lernen normalerweise effizient ist. Wenn ein Hinweisreiz mit hoher Vorhersagekraft existiert, werden ihn menschliche Beobachter finden, falls sie eine geeignete Gelegenheit dafür bekommen. In »verrauschten« Umgebungen sind statistische Algorithmen Menschen aus zwei Gründen überlegen: Sie spüren mit höherer Wahrscheinlichkeit schwach prädiktive Hinweisreize auf, und sie werden mit viel höherer Wahrscheinlichkeit auf lange Sicht mittelmäßig genaue Vorhersagen liefern, indem sie solche Hinweise konsequent verwerten.

Es ist falsch, jemandem einen Vorwurf daraus zu machen, dass er in einer unvorhersagbaren Welt keine genauen Vorhersagen liefert. Allerdings ist es durchaus angemessen, Experten dafür zu rügen, dass sie glauben, eine unmögliche Aufgabe erfolgreich bewältigen zu können. Die Behauptung, man habe in einer unvorhersagbaren Situation die richtigen Intuitionen gehabt, ist bestenfalls ein Selbstbetrug. Wenn prädiktive Hinweise fehlen, sind intuitive »Treffer« entweder auf Glück oder auf Lügen zurückzuführen. Wer diese Schlussfolgerung erstaunlich findet, glaubt noch immer, Intuition wäre eine magische Angelegenheit. Prägen Sie sich diese Regel ein: Wenn es keine stabilen Regelmäßigkeiten in der Umgebung gibt, kann man der Intuition nicht vertrauen.

## Feedback und Übung

Einige Regelmäßigkeiten in der Umgebung lassen sich leichter entdecken und anwenden als andere. Denken Sie daran, wie Sie Ihren Bremsstil beim Autofahren entwickelt haben. In dem Maße, wie Sie die Fähigkeit erwarben, Kurven zu fahren, lernten Sie, wann Sie vom Gas gehen und wann und wie fest Sie auf die Bremse treten sollten. Kurven unterscheiden sich, und die Variabilität, die Sie beim Lernen erlebten, stellt sicher, dass Sie jetzt bei jeder Kurve zur richtigen Zeit und mit der richtigen Stärke bremsen können. Die Bedingungen zum Erlernen dieser Fähigkeit sind ideal, weil Sie jedes Mal, wenn Sie durch eine Kurve fahren, sofort ein eindeutiges Feedback erhalten: die leichte Belohnung eines flüssigen Lenkmanövers oder die milde Bestrafung einer gewissen Schwierigkeit beim Lenken des Fahrzeugs, wenn man entweder zu scharf bremst oder nicht scharf genug. Die Situationen, mit denen ein Lotse beim Steuern großer Schiffe konfrontiert ist, sind nicht weniger regelmäßig, aber hier ist es wegen der langen Verzögerungen zwischen Manövern und ihren wahrnehmbaren Ergebnissen viel schwieriger, durch bloße Erfahrung Fertigkeiten zu erwerben. Ob Fachleute eine Chance haben, intuitive Expertise zu entwickeln, hängt im Wesentlichen von der Qualität und Schnelligkeit des Feedbacks sowie von ausreichenden Übungsgelegenheiten ab.

Expertise besteht nicht aus einer bestimmten Fähigkeit; sie ist eine Gesamtheit von Fähigkeiten, und derselbe Experte mag in einigen Aufgaben seines Fachgebiets sehr versiert sein, während er in anderen ein Neuling bleibt. Zu dem Zeitpunkt, zu dem Schachspieler Experten werden, haben sie »alles gesehen« (oder fast alles), aber Schach ist in dieser Hinsicht eine Ausnahme. Manche Chirurgen beherrschen gewisse Operationsverfahren viel besser als andere. Außerdem lassen sich einige Aspekte der Aufgaben jedes Experten viel leichter lernen. Psychotherapeuten haben viele Gelegenheiten, die unmittelbaren Reaktionen von Patienten auf ihre Äußerungen zu beobachten. Die Rückmeldung ermöglicht ihnen, die intuitive Fähigkeit zu entwickeln, die es ihnen erlaubt, die richtigen Worte und den richtigen Ton zu finden, um Wut zu lindern, Vertrauen aufzubauen oder die Aufmerksamkeit des

Patienten zu fokussieren. Andererseits haben Therapeuten keine Chance, herauszufinden, welches allgemeine therapeutische Verfahren für verschiedene Patienten am geeignetsten ist. Sie erhalten von ihren Patienten kaum oder nur mit großer zeitlicher Verzögerung oder auch (der Normalfall) gar keine Rückmeldungen über den langfristigen Therapieerfolg, und diese sind in jedem Fall nicht eindeutig genug, um ein Lernen aus Erfahrung zu ermöglichen.

Unter den Fachärzten profitieren die Anästhesisten von einem sachdienlichen Feedback, weil die Auswirkungen ihrer Maßnahmen in der Regel schnell deutlich werden. Radiologen dagegen erhalten kaum Informationen über die Richtigkeit ihrer Diagnosen und über die Erkrankungen, die sie nicht erkennen. Anästhesisten entwickeln daher eher nützliche intuitive Fähigkeiten. Wenn ein Anästhesist sagt: »Ich habe das Gefühl, dass etwas nicht stimmt«, sollte jeder im OP auf einen Notfall vorbereitet sein.

Auch hier mögen die Experten, wie im Fall der subjektiven Überzeugung, die Grenzen ihrer Expertise nicht erkennen. Eine erfahrene Psychotherapeutin weiß, dass sie recht zuverlässig einschätzen kann, was ihrem Patienten durch den Kopf geht, und dass sie intuitiv gut vorhersagen kann, was der Patient als Nächstes sagen wird. Die Annahme, auch antizipieren zu können, wie gut es dem Patienten im nächsten Jahr gehen wird, ist verlockend. Aber diese Schlussfolgerung ist nicht in gleicher Weise gerechtfertigt. Kurzfristige Antizipation und langfristige Vorhersage sind verschiedene Aufgaben, und die Therapeutin hatte hinlänglich Gelegenheit, das eine, nicht aber das andere zu lernen. In ähnlicher Weise mag ein Finanzexperte in vielen Aspekten seines Metiers sehr sachkundig sein, das gilt aber nicht für das Stock-Picking, und ein Nahost-Experte weiß vieles, kennt aber nicht die Zukunft. Der klinische Psychologe, der Stockpicker und der Experte verfügen bei einigen ihrer Aufgaben über intuitive Fähigkeiten, aber sie haben nicht gelernt, die Situationen und die Aufgaben zu identifizieren, bei denen die Intuition sie im Stich lässt. Die unerkannten Grenzen professioneller Sachkunde erklären, weshalb Experten ihre Fähigkeiten oftmals überschätzen.

## Die Gültigkeit von Intuitionen beurteilen

Am Ende unserer gemeinsamen Forschungen kamen Gary Klein und ich zu einer allgemeinen Antwort auf unsere Ausgangsfrage: Wann kann man einem erfahrenen Experten, der behauptet, eine zutreffende Intuition zu haben, vertrauen? Wir gelangten zu dem Schluss, dass es meistens möglich ist, zwischen wahrscheinlich zutreffenden und wahrscheinlich nicht zutreffenden Intuitionen zu unterscheiden. Wie bei der Beurteilung der Echtheit eines Kunstwerks ist es in der Regel sinnvoller, sich auf dessen Herkunft zu konzentrieren, statt das Kunstwerk selbst genauer zu prüfen. Wenn die Umgebung hinlänglich regelmäßig ist und wenn der Beurteiler die Gelegenheit hatte, ihre Regelmäßigkeiten zu erlernen, wird die Assoziationsmaschine Situationen wiedererkennen und schnell zutreffende Vorhersagen und Entscheidungen erzeugen. Wenn diese Bedingungen erfüllt sind, kann man sich auf die Intuitionen einer Person verlassen.

Leider erzeugt das assoziative Gedächtnis auch subjektiv hochgradig überzeugende Intuitionen, die falsch sind. Jeder, der die Fortschritte eines begabten jungen Schachspielers verfolgt hat, weiß nur zu gut, dass dieser nicht auf einmal zu einem perfekten Spieler wird und dass er auf dem Weg zur Beinahe-Perfektion höchstwahrscheinlich einige Fehler machen wird. Bei der Bewertung der intuitiven Urteile von Experten sollte man immer erwägen, ob der Experte hinlänglich Gelegenheit hatte, seine Fähigkeit zur Mustererkennung zu üben, auch in einem regelmäßigen Umfeld.

In einer weniger regelmäßigen Umgebung mit niedriger Vorhersagbarkeit nimmt man die Urteilsheuristiken zu Hilfe. System 1 kann oftmals durch Ersetzung rasche Antworten auf schwierige Fragen bereitstellen und Kohärenz erzeugen, wo keine vorhanden ist. Die Frage, die beantwortet wird, ist nicht die ursprüngliche Frage, aber die Antwort wird rasch erzeugt und mag hinreichend plausibel sein, um die lockere und nachsichtige Prüfung durch System 2 zu bestehen. Vielleicht wollen Sie den zukünftigen geschäftlichen Erfolg eines Unternehmens vorhersagen, und Sie glauben, dass Sie dies auch tun, während Ihre Beurteilung tatsächlich von Ihren Eindrücken über die Tatkraft und Kompetenz der gegenwärtigen Führungskräfte bestimmt wird. Weil die Ersetzung automatisch geschieht, kennen Sie oftmals nicht die Herkunft eines Urteils, das Sie (Ihr System 2)

gutheißen und übernehmen. Wenn es das Einzige ist, was Ihnen einfällt, mag es subjektiv nicht von gültigen Urteilen zu unterscheiden sein, die Sie mit der Sicherheit eines Experten treffen. Aus diesem Grund ist subjektives Überzeugtsein kein guter Maßstab für die Richtigkeit von Urteilen: Urteile, die die falsche Frage beantworten, können auch mit großer innerer Gewissheit getroffen werden.

Vielleicht fragen Sie sich, wieso Gary Klein und ich nicht sofort mit der Idee aufwarteten, die intuitiven Urteile eines Experten anhand der Regelmäßigkeit der Umgebung und der Lerngeschichte des Experten zu bewerten – unter weitgehender Ausklammerung der subjektiven Überzeugung des Experten. Und wie konnte unseres Erachtens die Antwort aussehen? Dies sind gute Fragen, weil sich die Lösung von Anfang an in Umrissen abzeichnete. Wir wussten von Beginn an, dass Einsatzleiter der Feuerwehr und Pädiatrie-Pflegekräfte schließlich auf der einen Seite der Grenze zutreffender intuitiver Urteile und dass die von Meehl untersuchten Fachgebiete zusammen mit den Stockpickern und den Wirtschaftsexperten auf der anderen Seite landen würden.

Es lässt sich heute schwer rekonstruieren, was dazu führte, dass wir Jahre brauchten, stundenlange Diskussionen, den endlosen Austausch von Entwürfen und Hunderte von E-Mails, in denen wir um Wörter feilschten, und mehr als einmal beinahe aufgegeben hätten. Aber dies geschieht immer, wenn ein Projekt einigermaßen gut zu Ende geht: Sobald man die zentrale Schlussfolgerung versteht, hat es den Anschein, sie wäre von jeher offensichtlich gewesen.

Wie aus dem Titel unseres Aufsatzes hervorgeht, waren Klein und ich uns nicht so uneinig, wie wir erwartet hatten, und wir akzeptierten für fast alle wichtigen Streitfragen, die angeschnitten wurden, gemeinsame Lösungen. Wir stellten jedoch auch fest, dass unsere frühen Differenzen mehr als eine intellektuelle Kontroverse waren. Wir hatten verschiedene Einstellungen, Emotionen und Geschmacksvorlieben, und diese veränderten sich im Lauf der Jahre bemerkenswert wenig. Dies zeigt sich am deutlichsten an den Tatsachen, die wir amüsant und interessant finden. Klein zuckt noch immer zusammen, wenn das Wort »Verzerrung« *(bias)* erwähnt wird, und er mag noch immer Geschichten, in denen Algorithmen oder formale Verfahren zu offensichtlich absurden Entscheidungen führen. Ich sehe in dem gelegentlichen Versagen von Algorithmen eher die Chance, sie zu verbessern. Andererseits finde ich mehr Vergnügen als Klein an der

wohlverdienten Strafe arroganter Experten, die sich in Situationen, die keine gültigen Vorhersagen erlauben, zutreffende intuitive Urteile anmaßen. Langfristig aber ist eine so große intellektuelle Übereinstimmung, wie wir sie gefunden haben, zweifellos wichtiger als die emotionalen Differenzen, die fortbestehen.

## Zum Thema »Intuition von Experten«

»Wie viel Expertise besitzt sie bei dieser konkreten Aufgabe? Wie viel Praxis hat sie?«

»Glaubt er wirklich, die Umgebung von Start-up-Firmen sei hinreichend regelmäßig, um ein intuitives Urteil zu rechtfertigen, das den Basisraten widerspricht?«

»Sie ist felsenfest von ihrer Entscheidung überzeugt, aber subjektives Überzeugtsein ist eine schlechte Kennzahl für die Genauigkeit eines Urteils.«

»Hatte er wirklich die Gelegenheit, zu lernen? Wie schnell und wie klar war das Feedback, das er über seine Urteile erhielt?«

# *23. Die Außensicht*

Wenige Jahre nach Beginn meiner Zusammenarbeit mit Amos überzeugte ich einige Beamte im israelischen Bildungsministerium von der Notwendigkeit, einen Lehrplan für den Unterricht in Urteils- und Entscheidungstheorie an Highschools zu erarbeiten. Das Team, das ich zusammenstellte, um das Curriculum zu entwickeln und ein Lehrbuch darüber zu schreiben, umfasste mehrere erfahrene Lehrer, einige meiner Psychologie-Studenten und Seymour Fox, damals Dekan der Erziehungswissenschaftlichen Fakultät der Hebräischen Universität, der ein Experte auf dem Gebiet der Curriculum-Entwicklung war.

Nachdem wir uns ein Jahr lang jeden Freitagnachmittag getroffen hatten, hatten wir eine detaillierte Kurzfassung des Syllabus (Lehrveranstaltungskonzepts) erarbeitet, ein paar Kapitel des Lehrbuchs geschrieben und einige Unterrichtsprobestunden gehalten. Wir alle hatten das Gefühl, gut voranzukommen. Als wir eines Tages Verfahren zur Abschätzung unbestimmter Größen diskutierten, kam mir die Idee, eine Übung durchzuführen. Ich bat alle Anwesenden, ihre Schätzung darüber aufzuschreiben, wie lange wir bräuchten, um einen fertigen Entwurf des Lehrbuchs beim Bildungsministerium einzureichen. Ich hielt mich dabei an eine Vorgehensweise, die wir bereits in unser Curriculum aufnehmen wollten: Die geeignete Methode, um Informationen von einer Gruppe zu gewinnen, besteht nicht darin, dass man eine öffentliche Diskussion beginnt, sondern darin, die Einschätzung jeder Person vertraulich einzusammeln. Dieses Verfahren nutzt das den Mitgliedern der Gruppe zur Verfügung stehende Wissen besser als die weitverbreitete Praxis der offenen Diskussion. Ich sammelte die Schätzwerte ein und schrieb die Ergebnisse an die Tafel. Die Punkte konzentrierten sich dicht um den Zeitraum von zwei Jahren; das untere Ende lag bei anderthalb, das obere bei zweieinhalb Jahren.

Dann hatte ich eine andere Idee. Ich wandte mich an Seymour, unseren Curriculum-Experten, und fragte ihn, ob ihm andere Teams einfielen, die ähnlich wie wir ein völlig neues Curriculum erarbeitet hatten. Damals waren gerade mehrere pädagogische Innovationen wie die »Neue

Mathematik« eingeführt worden, und Seymour sagte, ihm fielen etliche ein. Dann fragte ich, ob er Genaueres über die Geschichte dieser Teams wisse, und es stellte sich heraus, dass er einige persönlich kannte. Ich bat ihn, sich vorzustellen, diese Teams seien genauso weit wie wir. Wie lange bräuchten sie dann noch, um ihre Lehrbuchprojekte abzuschließen?

Er verstummte. Als er die Sprache endlich wiederfand, schien er zu erröten, durch seine eigene Antwort in Verlegenheit gebracht: »Wissen Sie, bislang hatte ich das noch gar nicht bemerkt, aber tatsächlich haben nicht alle Teams, die sich in einem ähnlichen Stadium wie wir befanden, ihre Vorhaben je zu Ende geführt. Ein erheblicher Prozentsatz der Teams hat das Projekt abgebrochen.«

Dies war beunruhigend; wir hatten nie die Möglichkeit in Betracht gezogen, zu scheitern. Mit wachsender Besorgnis fragte ich, wie hoch er diesen Prozentsatz schätze. »Etwa 40 Prozent«, antwortete er. Jetzt senkte sich ein Schleier der Düsternis über den Raum. Die nächste Frage lag auf der Hand: »Wie lange brauchten diejenigen, die durchhielten?« Er antwortete: »Mir fällt keine Gruppe ein, die in weniger als sieben Jahren fertig war, und keine, die länger als zehn Jahre brauchte.«

Ich griff nach dem Strohhalm: »Wenn Sie unsere Fähigkeiten und Ressourcen mit denen der anderen Gruppen vergleichen, wie gut sind wir dann? Wie schätzen Sie uns im Vergleich zu diesen Teams ein?« Diesmal zögerte Seymour nicht lange. »Wir sind unterdurchschnittlich«, sagte er, »aber nur ein bisschen.« Dies war für uns alle eine große Überraschung – auch Seymour selbst, dessen vorangehende Schätzung ganz klar innerhalb des optimistischen Gruppenkonsenses gelegen hatte. Erst als ich ihn dazu aufforderte, stellte er eine mentale Verbindung zwischen seiner Kenntnis der Erfolgsgeschichte anderer Teams und seiner Prognose unserer zukünftigen Leistung her.

Unser Geisteszustand, als wir Seymour hörten, lässt sich nicht gut mit dem Hinweis auf unser »Wissen« beschreiben. Sicherlich »wussten« wir alle, dass ein Minimum von sieben Jahren und eine 40-prozentige Wahrscheinlichkeit des Scheiterns eine plausiblere Prognose für das Schicksal unseres Projekts war als die Zahlen, die wir ein paar Minuten zuvor auf unsere Zettel geschrieben hatten. Aber wir erkannten nicht an, was wir wussten. Die neue Vorhersage wirkte noch immer irreal, weil wir uns nicht vorstellen konnten, wieso es so lange dauern sollte, ein Projekt abzuschließen, das so überschaubar zu sein schien. Wir konnten nicht in

eine Kristallkugel blicken, die uns die seltsame Abfolge unwahrscheinlicher Ereignisse vorausgesagt hätte, die uns bevorstanden. Alles, was wir sahen, war ein vernünftiger Plan, wonach wir in etwa zwei Jahren ein Buch produzieren sollten. Aber dieser Plan stand im Widerspruch zu statistischen Informationen, die besagten, dass andere Teams gescheitert waren beziehungsweise unglaublich lange gebraucht hatten, um ihr Projekt abzuschließen. Wir hatten Informationen über Basisraten erhalten, aus denen wir eine kausale Geschichte hätten ableiten sollen: Wenn so viele Teams scheiterten und wenn diejenigen, die erfolgreich waren, so lange brauchten, war das Schreiben eines Curriculums zweifellos viel schwieriger, als wir gedacht hatten. Aber eine solche Schlussfolgerung hätte unserer direkten Erfahrung der guten Fortschritte, die wir gemacht hatten, widersprochen. Die statistischen Daten, die Seymour bereitstellte, wurden so behandelt, wie Basisraten normalerweise behandelt werden – zur Kenntnis genommen und ad acta gelegt.

Wir hätten an diesem Tag aufgeben sollen. Niemand von uns war bereit, sechs weitere Arbeitsjahre in ein Projekt zu investieren, das mit einer Wahrscheinlichkeit von 40 Prozent scheitern würde. Obgleich wir gespürt haben müssen, dass ein Weitermachen nicht sinnvoll war, lieferte die Warnung keinen unmittelbar zwingenden Grund, aufzuhören. Nachdem wir ein paar Minuten halbherzig diskutiert hatten, nahmen wir uns zusammen und machten weiter, als ob nichts geschehen wäre. Das Buch wurde schließlich acht (!) Jahre später vollendet. Damals lebte ich nicht mehr in Israel und gehörte längst nicht mehr dem Team an, das die Aufgabe nach vielen unvorhergesehenen Widrigkeiten zu Ende brachte. Die anfängliche Begeisterung im Bildungsministerium für die Idee war zu dem Zeitpunkt, als das fertige Manuskript abgegeben wurde, längst verflogen.

Diese blamable Episode ist eine der lehrreichsten Erfahrungen meines Berufslebens geblieben. Ich lernte daraus schließlich drei Lektionen. Die erste war offensichtlich: Ich war auf eine Unterscheidung zwischen zwei grundverschiedenen Prognosemodellen gestoßen, die Amos und ich später als die »Innensicht« und die »Außensicht« bezeichneten.[1] Die zweite Lektion war, dass unsere anfänglichen Vorhersagen, wonach wir etwa zwei Jahre bräuchten, um das Projekt fertigzustellen, einen Planungsfehler aufwies. Unsere Schätzungen kamen einem optimalen Szenario näher als einer realistischen Beurteilung. Die dritte Lektion, die ich »irrationale Beharrlichkeit« nenne, habe ich nur langsam akzeptiert: die Unvernunft, die

wir an jenem Tag zeigten, als wir das Projekt nicht aufgaben. Vor die Wahl gestellt, ließen wir lieber die Vernunft sausen, als das Unternehmen aufzugeben.

## Die Verlockung der Innensicht

An jenem lange zurückliegenden Freitag gab unser Curriculum-Experte zwei Urteile über dasselbe Problem ab und gelangte zu sehr verschiedenen Antworten. [2] Die Innensicht ist diejenige, die sich alle von uns, einschließlich Seymour, spontan zu eigen machten, um die Zukunft unseres Projekts zu beurteilen. Wir konzentrierten uns auf unsere spezifischen Umstände und suchten nach Hinweisen in unseren eigenen Erfahrungen. Wir hatten einen groben Plan: Wir wussten, wie viele Kapitel wir schreiben wollten, und wir wussten, wie lange wir gebraucht hatten, um die beiden zu schreiben, die wir bereits fertiggestellt hatten. Die Vorsichtigeren unter uns schlugen vermutlich sicherheitshalber einige Monate auf ihren Schätzwert drauf.

Es war ein Fehler, einfach zu extrapolieren. Wir prognostizierten auf der Basis der Informationen, die uns vorlagen – gemäß der WYSIATI-Regel –, aber die zuerst geschriebenen Kapitel waren vermutlich einfacher als andere, und unser Engagement für das Projekt war damals am stärksten. Das Hauptproblem bestand jedoch darin, dass wir nicht – um es in Donald Rumsfelds Worten zu sagen – die »unbekannten Unbekannten« berücksichtigten. Wir konnten an diesem Tag nicht die Abfolge von Ereignissen vorhersehen, die dafür sorgte, dass sich das Projekt so in die Länge zog. Die Scheidungen, die Erkrankungen, die Abstimmungsprobleme mit Bürokratien, die die Arbeit verzögerten, waren nicht vorhersehbar. Solche Ereignisse führen nicht nur dazu, dass Kapitel langsamer fertiggestellt werden, sie erzeugen auch längere Phasen des Stillstands oder zäher Fortschritte. Das Gleiche muss natürlich für die anderen Teams gegolten haben, die Seymour kannte. Die Mitglieder dieser Teams konnten sich nicht die Ereignisse vorstellen, die die Ursache dafür waren, dass sie ein Projekt, das sie offenkundig für leicht realisierbar gehalten hatten, erst nach sieben Jahren beendeten oder unvollendet abbrachen. Wie wir, so wussten auch sie nichts über die Wahrscheinlichkeiten von Ereignissen. Pläne können auf vielfältige Weise scheitern, und obwohl die meisten dieser Möglichkeiten allzu unwahrscheinlich sind, als dass man sie vorhersehen könnte, ist die Wahrscheinlichkeit hoch, dass bei einem Großprojekt *irgendetwas* schiefgehen wird.

Die zweite Frage, die ich Seymour stellte, richtete seine Aufmerksamkeit von uns weg und auf eine Klasse ähnlicher Fälle. Seymour schätzte die Basisrate des Erfolgs in dieser Bezugsklasse: eine Versagensquote von 40 Prozent und sieben bis zehn Jahre bis zum Abschluss. Seine informelle Erhebung genügte gewiss nicht wissenschaftlichen Beweisstandards, aber sie lieferte eine vernünftige Grundlage für eine Basisprognose: die Prognose, die man über einen Fall anstellt, wenn man außer der Kategorie, zu der er gehört, keine weiteren Informationen besitzt. Wie wir früher sahen, sollte die Basisprognose der Anker für weitere Korrekturen sein. Wenn man Sie bittet, die Körpergröße einer Frau zu schätzen, von der Sie nur wissen, dass sie in New York City lebt, ist Ihre Basisprognose die beste Schätzung für die durchschnittliche Körpergröße von Frauen in dieser Stadt. Wenn Sie jetzt fallspezifische Informationen erhalten, zum Beispiel, dass der Sohn der Frau die Position des Starting Center in der Basketballmannschaft seiner Highschool innehat, werden Sie Ihre Schätzung in der geeigneten Richtung vom Mittelwert wegbewegen. Seymours Vergleich unseres Teams mit anderen deutete darauf hin, dass die Prognose unseres Ergebnisses etwas schlechter als die Basisprognose war, die bereits recht düster aussah.

Die sensationelle Genauigkeit der Außensichtprognose bei unserem Problem war mit Sicherheit ein Glückstreffer und sollte nicht als Beleg für die Gültigkeit der Außensicht betrachtet werden. Das Argument für die Außensicht sollte sich auf allgemeine Gründe stützen: Wenn die Bezugsklasse richtig ausgewählt wird, liefert die Außensicht einen Anhaltspunkt dafür, wo die »Hausnummer« zu suchen ist, und sie mag, wie in unserem Fall, darauf hindeuten, dass die Innensichtprognosen dieser nicht einmal nahekommen.

Für einen Psychologen ist die Diskrepanz zwischen den beiden Urteilen Seymours bemerkenswert. In seinem Kopf hatte er das gesamte Wissen, das notwendig war, um die statistischen Kenngrößen einer geeigneten Bezugsklasse abzuschätzen, aber er gelangte zu seinem anfänglichen Schätzwert, ohne dieses Wissen zu gebrauchen. Seymours Prognose aus seiner Innensicht war keine korrigierte Basisprognose, die ihm nicht eingefallen war. Sie basierte auf den besonderen Umständen unserer Bemühungen. Wie die Teilnehmer an dem Tom-W.-Experiment kannte Seymour die relevante Basisrate, dachte aber nicht daran, sie anzuwenden.

Anders als Seymour hatten wir Übrigen keinen Zugang zu der Außensicht und konnten daher keine realistische Basisprognose abgeben. Doch ist erwähnenswert, dass wir nicht das Gefühl hatten, Informationen über andere Teams zu brauchen, um unsere Schätzungen vorzunehmen. Meine Frage nach der Außensicht überraschte uns alle, mich eingeschlossen! Dies ist ein gängiges Muster: Menschen, die Informationen über einen Einzelfall besitzen, spüren nur selten das Bedürfnis, die statistischen Eckdaten der Klasse, zu der dieser Fall gehört, in Erfahrung zu bringen.

Als uns die Außensicht schließlich nahegebracht wurde, ignorierten wir sie. Wir können verstehen, was uns widerfuhr; es ist ganz ähnlich wie bei dem Experiment, das darauf hindeutet, dass die Vermittlung psychologischen Wissens nichts bringt. Als die Studenten von Nisbett und Borgida Vorhersagen über Einzelfälle machten, über die sie kaum Informationen besaßen (nur ein kurzes und nichtssagendes Interview), vernachlässigten sie vollständig die globalen Ergebnisse, die sie gerade gelernt hatten. »Blasse« statistische Informationen werden routinemäßig ausgesondert, wenn sie mit den persönlichen Eindrücken von einem Fall unvereinbar sind.[3] Im Wettstreit mit der Innensicht hat die Außensicht keine Chance.

Die Präferenz für die Innensicht ist manchmal mit moralischen Untertönen aufgeladen. Einmal stellte ich meinem Cousin, einem bekannten Anwalt, eine Frage nach der Bezugsklasse: »Wie hoch ist die Wahrscheinlichkeit, dass der Beklagte in Fällen wie diesen gewinnt?« Seine schroffe Antwort, »Jeder Fall ist einzigartig«, ging mit einem Gesichtsausdruck einher, der klarstellte, dass er meine Frage unangemessen und oberflächlich fand. Eine stolze Betonung der Einzigartigkeit der Fälle ist auch in der Medizin weit verbreitet, trotz der jüngsten Fortschritte in der evidenzbasierten Medizin, die in die andere Richtung weisen. Medizinische Statistiken und Basisprognosen fließen mit zunehmender Häufigkeit in Gespräche zwischen Patienten und Ärzten ein. Doch die fortbestehende Ambivalenz gegenüber der Außensicht in der Medizin drückt sich in Vorbehalten gegen die Unpersönlichkeit von Behandlungsverfahren aus, die sich auf Statistiken und Prüflisten stützen.[4]

## Der Planungsfehlschluss

In Anbetracht der Außensichtprognose und des letztendlichen Ergebnisses nehmen sich die ursprünglichen Schätzungen, die wir an jenem Freitagnachmittag machten, völlig unrealistisch aus. Dies sollte uns nicht weiter überraschen: Überoptimistische Vorhersagen der Ergebnisse von Projekten finden sich überall. Amos und ich prägten den Begriff »Planungsfehlschluss« *(planning fallacy)*, um Pläne und Prognosen zu beschreiben, die

– optimalen Szenarien unrealistisch nahekommen,
– verbessert werden könnten, wenn man die statistischen Daten über ähnlich gelagerte Fälle konsultierte.[5]

Privatpersonen, Behörden und Unternehmen sind mit einer Fülle von Planungsfehlschlüssen konfrontiert. Die Liste der Horrorgeschichten ist endlos.

– Im Juli 1997 wurden die Kosten für das neue Gebäude des schottischen Parlaments auf bis zu 40 Millionen Pfund geschätzt.[6] Im Juni 1999 betrug das Budget für den Neubau 109 Millionen Pfund. Im April 2000 beschlossen die Abgeordneten eine »Kostenobergrenze« von 195 Millionen Pfund. Im November 2001 verlangten sie eine Schätzung der »Endkosten«, die auf 241 Millionen Pfund beziffert wurden. Die geschätzten Endkosten wurden im Jahr 2002 zweimal nach oben korrigiert, um am Jahresende 294,6 Millionen Pfund zu erreichen. Im Jahr 2003 wurden sie drei weitere Male hinaufgesetzt und erreichten im Juni 375,8 Millionen Pfund. Das Gebäude wurde schließlich 2004 zu endgültigen Kosten von 431 Millionen Pfund fertiggestellt.
– Eine Studie aus dem Jahr 2005 nahm Eisenbahn-Neubaustrecken weltweit zwischen 1969 und 1998 unter die Lupe. In über 90 Prozent der Fälle wurde die Anzahl der Passagiere, die die Strecke nutzen würden, zu hoch veranschlagt. Obgleich diese überhöhten Prognosen weithin öffentlich kritisiert wurden, sind die Vorhersagen im Lauf dieser dreißig Jahre nicht genauer geworden; im Schnitt überschätzten die Planer die Passagierzahlen bei Neubauprojekten um 106 Prozent,

während die Durchschnittskosten um 45 Prozent über den vorhergesagten Kosten lagen. Obwohl immer mehr Daten zu Verfügung standen, haben sich die Experten nicht verstärkt darauf gestützt.[7]

– Im Jahr 2002 kam bei einer Studie über Eigenheimbesitzer in den USA, die ihre Küche renovierten, heraus, dass sie im Schnitt mit Kosten von 18 656 Dollar gerechnet hatten, während sie schließlich durchschnittlich 38 769 Dollar ausgaben.[8]

Der Optimismus von Planern und Entscheidern ist nicht die einzige Ursache von Kostenüberschreitungen. Firmen, die Küchen renovieren, und Rüstungsunternehmen geben durchaus zu (wenn auch nicht gegenüber ihren Kunden), dass sie den größten Teil ihres Gewinns mit Erweiterungen des ursprünglichen Auftrags verdienen. In diesen Fällen spiegelt sich in den Fehlprognosen die Unfähigkeit der Kunden wider, sich vorzustellen, dass sie im Lauf der Zeit immer neue Wünsche entwickeln. Schließlich bezahlen sie viel mehr, als wenn sie einen realistischen Plan erstellt und sich daran gehalten hätten.

Fehler in der anfänglichen Budgetierung sind nicht immer schuldlos. Diejenigen, die unrealistische Pläne aufstellen, sind oftmals von dem Wunsch getrieben, den Plan bewilligt zu bekommen – egal ob von ihren Vorgesetzten oder von einem Kunden. Hinzu kommt das Wissen, dass Projekte nur selten wegen Kosten- oder Terminüberschreitungen halb fertig aufgegeben werden.[9] In diesen Fällen liegt die Hauptverantwortung für die Vermeidung des Planungsfehlschlusses bei den Entscheidern, die den Plan bewilligen. Wenn sie nicht die Notwendigkeit einer Außensicht erkennen, begehen sie einen Planungsfehlschluss.

## Wie man Planungsfehlschlüsse in den Griff bekommt

Diagnose und Gegenmittel des Planungsfehlschlusses haben sich seit jenem Freitagnachmittag nicht verändert, aber bei der Umsetzung des Konzepts wurden große Fortschritte gemacht. Der angesehene dänische Planungsexperte Bent Flyvbjerg, der heute an der Universität Oxford lehrt, lieferte eine eindrucksvolle Zusammenfassung:

> Die vorherrschende Tendenz, verteilte Informationen unterzugewichten oder zu ignorieren, ist vielleicht die Hauptursache von Fehlprognosen. Planer sollten sich daher intensiv darum bemühen, das Vorhersageproblem so zu formulieren, dass die Nutzung sämtlicher verteilter Informationen, die verfügbar sind, erleichtert wird.

Dies ist vielleicht der wichtigste Ratschlag, den man geben kann, um durch verbesserte Methoden die Genauigkeit von Vorhersagen zu steigern. Die Nutzung verteilter Informationen aus anderen Projekten, die vergleichbar sind mit dem, über das eine Prognose erstellt wird, entspricht einer Außensicht und ist die geeignete Abhilfe für den Planungsfehlschluss.

Der Fachbegriff für die Methode zur Beseitigung des Planungsfehlschlusses lautet »Referenzklassenprognose«, und Flyvberg hat diese Methode auf Verkehrsprojekte in mehreren Ländern angewandt. Die Implementierung der Außensicht erfolgt mithilfe einer großen Datenbank, in der für Hunderte von Projekten weltweit Informationen über Pläne und die tatsächlichen Ergebnisse gespeichert sind, und diese Außensicht liefert statistische Informationen über wahrscheinliche Kosten- und Terminüberschreitungen sowie darüber, mit welcher Wahrscheinlichkeit verschiedenste Projekte ihre Zielvorgaben verfehlen.

Die von Flyvberg angewandte Prognosemethode ähnelt den Handlungsempfehlungen, mit denen sich die Vernachlässigung der Basisrate verhindern lässt:

1. Man identifiziere eine geeignete Referenzklasse (Küchenrenovierungen, große Eisenbahn-Neubauprojekte usw.).
2. Man beschaffe sich die statistischen Eckdaten der Referenzklasse (in Bezug auf Kosten pro Schienenkilometer oder den Prozentsatz, um den die Ausgaben den Kostenrahmen überschreiten). Mithilfe dieser statistischen Daten erzeuge man eine Basisprognose.
3. Man benutze die spezifischen Informationen über den Einzelfall, um die Basisprognose zu korrigieren, wenn es konkrete Gründe für die Erwartung gibt, dass die optimistische Verzerrung bei diesem Projekt stärker oder weniger stark ausgeprägt ist als bei anderen Projekten des gleichen Typs.

Flyvbergs Analysen sind als Leitfaden für Behörden gedacht, die öffentliche Aufträge vergeben, indem sie statistische Daten zu Überschreitungen bei ähnlichen Projekten bereitstellen. Entscheidungsträger brauchen eine realistische Einschätzung der Kosten und des Nutzens eines Projekts, bevor sie die endgültige Bewilligungsentscheidung treffen. Vielleicht möchten sie auch die Budgetreserve abschätzen, die sie in Erwartung von Überschreitungen brauchen, obwohl solche Vorsichtsmaßnahmen oftmals zu sich selbst erfüllenden Prophezeiungen werden. So äußerte ein Beamter gegenüber Flyvberg: »Eine Budgetreserve ist für Auftragnehmer das, was rotes Fleisch für Löwen ist: Sie verschlingen sie.«

Organisationen sind mit der Herausforderung konfrontiert, die Neigung ihrer Führungskräfte in Schach zu halten, im Konkurrenzkampf um Ressourcen übermäßig optimistische Pläne vorzulegen. Eine gut geführte Organisation belohnt Planer für eine präzise Ausführung und bestraft sie, wenn sie vorhersehbare Schwierigkeiten nicht vorhersehen und wenn sie nicht vorhersehbare Schwierigkeiten nicht einkalkulieren – die unbekannten Unbekannten.

## Entscheidungen und Irrtümer

Dieser Freitagnachmittag liegt jetzt über dreißig Jahre zurück. Ich habe oft darüber nachgedacht, und ich habe ihn in meinen Vorlesungen mehrmals jährlich erwähnt. Einige meiner Freunde wurden dieser Geschichte überdrüssig, aber ich habe immer wieder neue Lektionen daraus gelernt. Fast 15 Jahre nachdem ich zum ersten Mal gemeinsam mit Amos über den Planungsfehlschluss publiziert hatte, kehrte ich mit Dan Lovallo zu dem Thema zurück. Zusammen erarbeiteten wir die Grundzüge einer Theorie der Entscheidungsfindung, in der die Optimismus-Verzerrung eine wichtige Ursache der Risikobereitschaft darstellt. Im rationalen Standardmodell der Volkswirtschaftslehre gehen Menschen deshalb Risiken ein, weil die Chancen günstig sind – sie nehmen eine gewisse Wahrscheinlichkeit für einen kostspieligen Fehlschlag in Kauf, weil die Erfolgswahrscheinlichkeit hinlänglich hoch ist. Wir schlugen ein alternatives Modell vor.

Bei der Prognose von Ergebnissen risikoreicher Projekte fallen Führungskräfte allzu leicht dem Planungsfehlschluss zum Opfer. Unter seinem Einfluss treffen sie Entscheidungen, die auf irrationalem Optimismus statt auf einer rationalen Abwägung von Gewinnen, Verlusten und Wahrscheinlichkeiten basieren. Sie überschätzen den Nutzen und unterschätzen die Kosten. Sie denken sich Erfolgsszenarien aus, während sie das Potenzial für Fehler und Fehlberechnungen übersehen. Aus diesem Grund verfolgen sie Projekte, bei denen es eher unwahrscheinlich ist, dass sie sich innerhalb des Kosten- oder Zeitrahmens bewegen oder die erwarteten Erträge abwerfen werden – oder auch nur zur Vollendung kommen.

Nach dieser Sichtweise lassen sich Menschen oft (aber nicht immer) auf riskante Projekte ein, weil sie deren Erfolgsaussichten allzu optimistisch einschätzen. Ich werde im weiteren Verlauf dieses Buches noch einige Male auf diese Idee zurückkommen – sie erklärt vermutlich mit, weshalb Menschen prozessieren, weshalb sie Kriege anzetteln und kleine Firmen gründen.

## Einen Test nicht bestehen

Viele Jahre lang glaubte ich, der entscheidende Punkt der Curriculum-Geschichte sei das, was ich über meinen Freund Seymour gelernt hatte – dass sich seine Einschätzung der Zukunft unseres Projekts nicht auf das stützte, was er über ähnliche Projekte wusste. In meiner Version der Geschichte, in der ich die Rolle des klugen Fragenstellers und des scharfsinnigen Psychologen spielte, kam ich ziemlich gut weg. Erst vor Kurzem wurde mir klar, dass ich in Wirklichkeit die Rollen des obersten Dummkopfs und des unfähigen Anführers gespielt hatte.

Das Projekt ging auf meine Initiative zurück, und aus diesem Grund oblag es mir, dafür zu sorgen, es zu aussagekräftigen Ergebnissen zu führen und wichtige Probleme vom Team sachgerecht diskutieren zu lassen. Aber ich habe diesen Test nicht bestanden. Mein Problem war nicht länger der Planungsfehlschluss. Sobald ich Seymours Zusammenfassung der statistischen Daten hörte, war ich von diesem Fehlschluss geheilt. Hätte man mich bedrängt, hätte ich gesagt, dass unsere früheren Schätzungen in geradezu absurder Weise optimistisch gewesen seien. Hätte man mich noch stärker bedrängt, dann hätte ich eingeräumt, dass wir das Projekt auf der Basis falscher Prämissen begonnen hatten und dass wir die Option, uns geschlagen zu geben und nach Hause zu gehen, wenigstens in Erwägung ziehen sollten. Aber niemand bedrängte mich, und es gab keine Diskussion; wir kamen stillschweigend überein, ohne explizite Prognose über die Dauer des Projekts weiterzumachen. Dies war einfach, weil wir am Anfang keine derartige Prognose erstellt hatten. Hätten wir schon zu Beginn eine vernünftige Basisprognose gehabt, dann hätten wir das Projekt gar nicht erst angepackt, aber jetzt hatten wir bereits viel Arbeit investiert – ein Beispiel für den Fehlschluss der »versunkenen (irreversiblen) Kosten«, mit dem wir uns im nächsten Teil des Buches eingehender befassen werden.[10] Es wäre für uns – insbesondere für mich – blamabel gewesen, an diesem Punkt aufzugeben, und es schien keinen direkten Grund dafür zu geben. In einer Krise ist es einfacher, die Richtung zu wechseln, aber dies war keine Krise, vielmehr erfuhren wir nur ein paar neue Fakten über Menschen, die wir nicht kannten. Die Außenperspektive ließ sich viel leichter ausblenden als schlechte Neuigkeiten, auf die wir selbst stießen. Unser Zustand lässt sich

am besten als eine Art Lethargie beschreiben – die mangelnde Bereitschaft, über das nachzudenken, was passiert war. Also machten wir weiter. Während der restlichen Zeit, die ich als Mitglied des Teams verbrachte, gab es keinen weiteren Versuch zur rationalen Planung – ein besonders beunruhigendes Versäumnis für ein Team, das rationales Denken lehren wollte. Ich hoffe, dass ich heute klüger bin, und ich habe es mir zur Gewohnheit gemacht, die Außensicht zu berücksichtigen. Aber das wird nie selbstverständlich für mich sein.

## Zum Thema »Außensicht«

»Er betrachtet die Dinge aus der Innenperspektive. Er sollte seinen eigenen Fall vergessen und sich ansehen, was in anderen Fällen geschehen ist.«

»Sie ist das Opfer eines einfachen Planungsfehlschlusses. Sie geht von einem optimalen Szenario aus, aber der Plan kann auf sehr viele unterschiedliche Weisen scheitern, und sie kann diese Möglichkeiten nicht alle vorhersehen.«

»Angenommen, Sie wüssten nichts über diesen konkreten Prozess, nur dass es dabei um eine Schadenersatzforderung wegen eines angeblichen chirurgischen Kunstfehlers geht. Wie sähe Ihre Basisprognose aus? Wie viele dieser Verfahren enden mit einem Sieg für den Kläger? Wie viele werden durch einen Vergleich beendet? Um welche Summen geht es dabei? Ist der Fall, über den wir diskutieren, stärker oder schwächer als vergleichbare Verfahren?«

»Wir tätigen eine zusätzliche Investition, weil wir nicht zugeben wollen, dass wir gescheitert sind. Dies ist ein Beispiel für den Fehlschluss der versunkenen Kosten.«

# *24. Die Maschine des Kapitalismus*

Der Planungsfehlschluss ist nur eine der Erscheinungsformen einer weitverbreiteten optimistischen Verzerrung. Die meisten von uns sehen die Welt positiver, als sie es tatsächlich ist, sich selbst in einem günstigeren Licht, als es angemessen ist, und die Ziele, die sie sich vornehmen, empfinden sie als leichter realisierbar, als sie es tatsächlich sind. Wir neigen auch dazu, unsere Fähigkeit überzubewerten, die Zukunft vorherzusagen, was eine optimistische Selbstüberschätzung begünstigt. Im Hinblick auf ihre Folgen für unsere Entscheidungen mag die Optimismus-Verzerrung durchaus die wichtigste kognitive Verzerrung sein. Weil die Optimismus-Verzerrung sowohl ein Segen als auch ein Fluch sein kann, sollten Sie sich einerseits glücklich schätzen, aber andererseits auf der Hut sein, wenn Sie von Natur aus optimistisch sind.

## Optimisten

Optimismus ist normal, aber einige glückliche Menschen sind optimistischer als wir Übrigen. Wenn Sie genetisch mit einer optimistischen Einstellung ausgestattet wurden, braucht man Ihnen kaum zu sagen, dass Sie sich glücklich schätzen können – Sie fühlen sich bereits vom Glück begünstigt.[1] Eine optimistische Einstellung ist größtenteils erblich bedingt, und sie ist Teil einer allgemeinen Disposition zum Wohlbefinden, die auch die Tendenz umfassen mag, bei allem die positive Seite zu sehen.[2] Wenn Sie für Ihr Kind einen Wunsch frei hätten, sollten Sie ernsthaft in Betracht ziehen, ihm Optimismus zu wünschen. Optimisten sind normalerweise fröhlich und zufrieden und daher beliebt; sie kommen gut mit Fehlschlägen und Notlagen zurecht, sie haben ein geringeres Risiko, an einer klinischen Depression zu erkranken, ihr Immunsystem ist stärker, sie achten besser auf ihre Gesundheit, sie fühlen sich gesünder als andere und sie haben tatsächlich eine höhere Lebenserwartung. Eine Studie über Menschen, die davon überzeugt sind, dass ihre Lebenserwartung über den statistischen Vorhersagen der Versicherungen liegt, zeigte, dass sie länger arbeiten, ihr zukünftiges Einkommen optimistischer einschätzen und nach einer Scheidung mit höherer Wahrscheinlichkeit erneut heiraten (der klassische »Triumph der Hoffnung über die Erfahrung«) sowie eher dazu neigen, auf einzelne Aktien zu setzen.[3] Selbstverständlich profitieren nur Menschen mit einer leichten rosaroten Brille von den Segnungen des Optimismus; sie sind in der Lage, »das Positive zu akzentuieren«, ohne die Wirklichkeit aus den Augen zu verlieren.

Optimistische Menschen üben einen unverhältnismäßig starken Einfluss auf unser Leben aus. Ihre Entscheidungen haben weitreichende Folgen; sie sind die Erfinder, die Unternehmer, die politischen und militärischen Führungsfiguren – keine gewöhnlichen Menschen. Sie gelangten in ihre Positionen, weil sie Herausforderungen suchten und Risiken eingingen. Sie sind begabt, und sie hatten Glück, höchstwahrscheinlich mehr Glück, als sie zugeben. Sie sind vermutlich geborene Optimisten; bei einer Studie über die Gründer von kleinen Unternehmen kam heraus, dass Unternehmer dem Leben allgemein zuversichtlicher gegenüberstehen als mittlere

Führungskräfte.[4] Ihre Erfolgserlebnisse haben ihren Glauben an ihre Urteilskraft und an ihre Fähigkeit, Ereignisse zu kontrollieren, bestätigt. Ihr Selbstbewusstsein wird durch die Bewunderung anderer verstärkt.[5] Diese Überlegung führt zu einer Hypothese: Die Menschen, die den größten Einfluss auf das Leben anderer ausüben, sind vermutlich optimistisch und übertrieben selbstbewusst, und sie gehen mehr Risiken ein, als ihnen klar ist.[6]

Die empirischen Daten deuten darauf hin, dass eine Optimismus-Verzerrung immer dann eine Rolle spielt – manchmal die entscheidende Rolle –, wenn Personen oder Institutionen freiwillig erhebliche Risiken eingehen. Meistens unterschätzen risikofreudige Akteure die Risiken, die sie eingehen, und sie bemühen sich nicht hinlänglich, die Höhe der Risiken herauszufinden. Weil sie die Risiken falsch einschätzen, halten sich optimistische Unternehmer oftmals für besonnen, auch wenn sie dies nicht sind. Ihr Glaube an ihren zukünftigen Erfolg fördert eine positive Stimmung, die ihnen hilft, sich Finanzmittel von anderen zu beschaffen, die Arbeitsmoral ihrer Beschäftigten zu heben und die Erfolgschancen ihrer Firma zu erhöhen. Wenn Tatkraft gefragt ist, ist Optimismus, selbst wenn er leicht wahnhafte Züge trägt, eine gute Sache.

## Illusionen von Unternehmern

Die Wahrscheinlichkeit, dass ein neu gegründetes Kleinunternehmen in den Vereinigten Staaten die ersten fünf Jahre überstehen wird, beträgt etwa 35 Prozent. Aber die Personen, die solche Unternehmen gründen, glauben nicht, dass diese statistischen Werte für sie gelten. Laut einer Studie neigen amerikanische Unternehmer zu der Ansicht, sie seien in einer vielversprechenden Branche tätig: Sie schätzten die Erfolgsaussichten für »jedes Unternehmen wie das eigene« im Schnitt auf 60 Prozent – fast das Doppelte des wahren Wertes. Die Verzerrung war noch eklatanter, wenn die Befragten die Erfolgsaussichten ihres eigenen Projekts beurteilen sollten. Volle 81 Prozent der Unternehmer bezifferten ihre Erfolgsaussichten auf mindestens 70 Prozent, und 33 Prozent sagten, ihr Risiko zu scheitern sei gleich null.[7]

Die Richtung der Verzerrung ist nicht weiter überraschend. Wenn Sie jemanden befragen würden, der vor Kurzem ein italienisches Restaurant eröffnete, würden Sie nicht erwarten, dass diese Person ihre Erfolgsaussichten unterschätzen oder ihre Fähigkeiten als Gastronom gering achten würde. Aber Sie müssen sich fragen: Hätte diese Person auch dann Zeit und Geld investiert, wenn sie sich hinlänglich angestrengt hätte, die statistische Erfolgswahrscheinlichkeit in Erfahrung zu bringen – beziehungsweise, hätte sie, wenn sie die statistische Wahrscheinlichkeit herausgefunden hätte (60 Prozent aller neuen Gaststätten machen nach drei Jahren wieder zu), dieser Beachtung geschenkt? Vermutlich kam sie nicht auf den Gedanken, sich die Außensicht zu eigen zu machen.

Einer der Vorteile eines optimistischen Temperaments besteht darin, dass es Beharrlichkeit trotz Hindernissen fördert. Aber Beharrlichkeit kann kostspielig sein. Eine eindrucksvolle Reihe von Studien von Thomas Åstebro gibt Aufschluss darüber, was geschieht, wenn Optimisten schlechte Nachrichten erhalten. Seine Daten stammten von einer kanadischen Organisation – dem Inventor’s Assistance Program –, die Erfindern gegen eine kleine Gebühr eine objektive Beurteilung der gewerblichen Erfolgsaussichten ihrer Idee anbietet. Die Beurteilungen stützen sich auf eine sorgfältige Bewertung jeder Erfindung nach 37 Kriterien, wie etwa die Frage, ob ein Bedarf an dem Produkt besteht, die Produktionskosten und

der geschätzte Nachfragetrend. Die Analytiker fassen ihre Bewertungen in einer Buchstabennote zusammen, wobei die Buchstaben D und E ein Scheitern vorhersagen – eine Vorhersage, die für über 70 Prozent der Erfindungen gemacht wird, die sie unter die Lupe nehmen. Die Treffgenauigkeit der Misserfolgs-Prognosen ist erstaunlich hoch: Nur 5 von 411 Projekten, die die niedrigste Bewertung erhielten, wurden kommerzialisiert, und keines davon war erfolgreich.[8]

Entmutigende Nachrichten veranlassten etwa die Hälfte der Erfinder dazu, ihre Projekte nicht weiterzuverfolgen, nachdem diese eine Note erhalten hatten, die unmissverständlich ein Scheitern vorhersagte. Aber 47 Prozent der Erfinder setzten ihre Entwicklungsbemühungen auch dann noch fort, nachdem ihnen die Aussichtslosigkeit ihres Projekts mitgeteilt worden war, und im Schnitt haben diese beharrlichen (beziehungsweise starrsinnigen) Personen ihre anfänglichen Verluste verdoppelt, ehe sie aufgaben. Bezeichnenderweise zeigte sich diese Beharrlichkeit nach einem entmutigenden Rat vor allem bei Erfindern, die in der Persönlichkeitsdimension Optimismus einen hohen Punktwert erzielten – in der Erfinder allgemein schon höhere Werte verzeichnen als der Bevölkerungsdurchschnitt. Insgesamt war der Kapitalertrag aus privaten Erfindungen gering, »geringer als die Rendite von Private-Equity-Anlagen oder hochriskanten Wertpapieren«. Ganz allgemein gilt, dass die finanziellen Vorteile einer selbstständigen Erwerbstätigkeit eher mäßig sind: Bei gleicher Qualifikation ist es für Erwerbstätige finanziell vorteilhafter, ihre Kompetenzen an einen Arbeitgeber zu verkaufen, als sich damit selbstständig zu machen. Die empirischen Daten deuten darauf hin, dass Optimismus weit verbreitet, hartnäckig und kostspielig ist.[9]

Psychologen haben bestätigt, dass die meisten Menschen ernsthaft glauben, sie seien in den meisten wünschenswerten Persönlichkeitszügen der Mehrheit überlegen – bei Laborversuchen sind sie bereit, kleine Summen Geld auf diese Überzeugungen zu setzen.[10] Am Markt hat der Glaube an die eigene Überlegenheit natürlich weitreichende Folgen. Die Chefs großer Unternehmen schließen manchmal gewaltige Wetten auf teure Fusionen und Akquisitionen ab, wobei sie von der falschen Annahme ausgehen, sie könnten die Vermögenswerte eines anderen Unternehmens besser nutzen als dessen gegenwärtige Eigentümer. An der Börse kommt es daraufhin in der Regel zu einem Kurseinbruch bei der Aktie der

übernehmenden Gesellschaft, weil die Erfahrung gezeigt hat, dass Bemühungen zur Integration großer Firmen meistens scheitern. Diese verfehlten Akquisitionen wurden mit der »Hybris-Hypothese« erklärt: Die Führungskräfte der übernehmenden Gesellschaft sind schlichtweg weniger kompetent, als sie zu sein glauben.[11]

Die Ökonomen Ulrike Malmendier und Geoffrey Tate identifizierten optimistische Vorstandschefs anhand der Anzahl von Firmenaktien, die sie persönlich hielten, und sie beobachteten, dass hochoptimistische Führungskräfte unverhältnismäßig hohe Risiken eingingen. Sie nahmen lieber Fremdkapital auf, als Aktien auszugeben, und sie waren eher bereit als andere, »für Zielgesellschaften einen überhöhten Preis zu zahlen und wertvernichtende Fusionen durchzuführen«.[12] Bemerkenswerterweise erlebte die Aktie der übernehmenden Gesellschaft einen viel stärkeren Kurseinbruch, wenn der Vorstandschef nach dem Maßstab der Autoren allzu optimistisch war. Offensichtlich ist die Börse in der Lage, übertrieben zuversichtliche Vorstandschefs zu erkennen. Diese Beobachtung entlastet die Vorstandschefs von einem Vorwurf, auch wenn sie sie in einem anderen Punkt schuldig spricht: Unternehmensführer, die leichtfertige Wetten eingehen, tun dies nicht, weil sie mit dem Geld anderer Leute wetten. Im Gegenteil, sie gehen höhere Risiken ein, wenn für sie persönlich mehr auf dem Spiel steht. Der Schaden, den überoptimistische Vorstandschefs anrichten, wird noch verschlimmert, wenn die Wirtschaftspresse ihnen zu Ruhm und Ansehen verhilft; vieles spricht dafür, dass renommierte Auszeichnungen von Spitzenmanagern durch die Presse die Aktionäre teuer zu stehen kommen. Die Autoren schreiben: »Wir haben herausgefunden, dass sich Unternehmen mit preisgekrönten Chefs im Anschluss an die Auszeichnung unterdurchschnittlich entwickeln, sowohl was ihren Aktienkurs anlangt als auch im Hinblick auf ihre operative Leistungsfähigkeit. Gleichzeitig erhöht sich die Vergütung der Vorstandschefs, die jetzt mehr Zeit auf Aktivitäten außerhalb des Unternehmens verwenden, wie etwa das Schreiben von Büchern oder die Wahrnehmung externer Verwaltungsratspositionen, und sie betreiben jetzt auch eher Bilanzkosmetik.«[13]

Vor vielen Jahren machten meine Frau und ich Urlaub auf Vancouver Island, wo wir nach einer geeigneten Unterkunft Ausschau hielten. Wir

fanden ein hübsches, aber menschenleeres Motel an einer kaum befahrenen Straße mitten in einem Wald. Es gehörte einem bezaubernden jungen Paar, das uns bereitwillig seine Geschichte erzählte. Sie waren Lehrer in der Provinz Alberta gewesen; sie hatten beschlossen, einen Neuanfang zu wagen, und mit ihren Ersparnissen kauften sie dieses Motel, das rund zehn Jahre zuvor erbaut worden war. Sie erzählten uns ohne Ironie oder Verlegenheit, dass sie es billig kaufen konnten, weil »sechs oder sieben vorhergehende Eigentümer nichts daraus gemacht haben«. Sie erzählten uns auch von Plänen, einen Kredit aufzunehmen, um die Anlage attraktiver zu gestalten, indem sie gleich daneben ein Restaurant eröffneten. Sie hielten es nicht für notwendig, zu erklären, weshalb sie erwarteten, erfolgreich zu sein, wo sechs oder sieben andere gescheitert waren. Ein gemeinsamer Faden von Wagemut und Optimismus verbindet Geschäftsleute, von Motelbesitzern bis zu Superstar-Wirtschaftskapitänen.

Die optimistische Risikobereitschaft von Unternehmern trägt zweifellos zu der wirtschaftlichen Dynamik einer kapitalistischen Gesellschaft bei, auch wenn die meisten risikobereiten Unternehmer schließlich enttäuscht sind. Doch Marta Coelho von der London School of Economics hat auf die heiklen politischen Probleme hingewiesen, die entstehen, wenn Gründer mittelständischer Unternehmen Behörden bitten, sie bei Entscheidungen zu unterstützen, die sich mit hoher Wahrscheinlichkeit als Fehlschläge erweisen werden. Sollte der Staat Kredite für Möchtegernunternehmer bereitstellen, die vermutlich in ein paar Jahren pleitegehen werden? Viele Verhaltensökonomen sind voll und ganz einverstanden mit den »libertär-paternalistischen« Vorgehensweisen, die Menschen dazu bringen, ihre Sparquote über das Maß hinaus zu steigern, das sie von sich aus als Grenze setzen würden. Die Frage, ob und wie der Staat die mittelständische Wirtschaft unterstützen sollte, hat keine vergleichbar befriedigende Antwort.

## Vernachlässigung der Konkurrenz

Es ist verlockend, unternehmerischen Optimismus als Wunschdenken zu erklären, aber Emotionen sind nur ein Teil der Geschichte. Kognitive Verzerrungen spielen eine wichtige Rolle, insbesondere die Eigenschaft von System 1, nur die aktuell verfügbaren Informationen zu berücksichtigen.

- Wir konzentrieren uns auf unser Ziel, verankern uns in unserem Plan und vernachlässigen relevante Basisraten, sodass wir uns dem Planungsfehlschluss aussetzen.
- Wir konzentrieren uns auf das, was wir tun wollen und können, und vernachlässigen die Pläne und Fähigkeiten von anderen.
- Sowohl bei der Erklärung vergangener Ereignisse als auch bei der Vorhersage der Zukunft konzentrieren wir uns auf die kausale Rolle von Fähigkeiten und vernachlässigen die Rolle des Zufalls. Daher sind wir anfällig für die Illusion der Kontrolle.
- Wir konzentrieren uns auf das, was wir wissen, und vernachlässigen das, was wir nicht wissen, sodass wir die Richtigkeit unserer Überzeugungen überschätzen.

Die Beobachtung, dass »90 Prozent der Fahrer glauben, überdurchschnittlich gut zu fahren«, ist ein gut gesicherter psychologischer Befund, der Gemeingut geworden ist, und er wird häufig als ein vorzügliches Beispiel für einen allgemeineren Above-Average-Effekt (die Tendenz, sich selbst eine überdurchschnittliche Leistungsfähigkeit zuzuschreiben) betrachtet. Doch die Interpretation dieses Befundes hat sich in den letzten Jahren gewandelt, von einer Selbstwerterhöhung zu einer kognitiven Verzerrung.[14] Betrachten Sie diese beiden Fragen:

Sind Sie ein guter Fahrer?
Sind Sie ein überdurchschnittlich guter Fahrer?

Die erste Frage ist einfach, und die Antwort erfolgt schnell: Die meisten Fahrer sagen Ja. Die zweite Frage ist viel schwieriger, und für die meisten Befragten ist es beinahe unmöglich, sie richtig zu beantworten, weil sie eine Beurteilung der durchschnittlichen Qualität von Fahrern erfordert. An dieser Stelle des Buches überrascht es uns nicht mehr, dass Menschen eine schwierige Frage dadurch beantworten, dass sie eine leichtere lösen. Sie vergleichen sich mit dem Durchschnitt, ohne jemals über den Durchschnitt nachzudenken. Der Beleg für die kognitive Interpretation des Above-Average-Effekts besteht darin, dass Menschen, wenn sie zu einer Aufgabe befragt werden, die ihnen schwierig erscheint (für viele von uns zum Beispiel: »Sind Sie überdurchschnittlich gut, wenn es darum geht, Gespräche mit Fremden anzuknüpfen?«), sich schnell als unterdurchschnittlich einstufen. Dies führt letzten Endes dazu, dass Menschen ihre relative Leistungsfähigkeit in jeder Aktivität, in der sie einigermaßen gut sind, übermäßig optimistisch einschätzen.

Ich hatte mehrfach Gelegenheit, Firmengründern und Teilhabern innovativer Start-ups eine Frage zu stellen: In welchem Ausmaß hängt das Ergebnis Ihrer Bemühungen davon ab, was Sie in Ihrem Unternehmen tun? Dies ist offensichtlich eine leichte Frage; die Antwort kommt schnell, und in meiner kleinen Stichprobe lag sie nie unter 80 Prozent. Selbst wenn sie nicht sicher sind, ob sie erfolgreich sein werden, glauben diese wagemutigen Personen, ihr Schicksal liege fast zur Gänze in ihrer Hand. Sie irren sich zweifellos: Der Erfolg eines Start-up-Unternehmens hängt ebenso sehr von den Leistungen ihrer Wettbewerber und von Marktentwicklungen wie von ihren eigenen Anstrengungen ab. Doch auch die WYSIATI-Regel spielt eine Rolle, und Unternehmer konzentrieren sich automatisch auf das, was sie am besten wissen – ihre Pläne und Handlungen und die unmittelbarsten Bedrohungen und Chancen, wie etwa die Verfügbarkeit von Kapital. Sie wissen weniger über ihre Konkurrenten, und sie stellen sich daher ganz selbstverständlich eine Zukunft vor, in der die Konkurrenten keine große Rolle spielen.

Colin Camerer und Dan Lovallo, die den Begriff der »Konkurrenzvernachlässigung« prägten, verdeutlichten dies mit einem Zitat des damaligen Chefs der Disney Studios. Danach gefragt, weshalb so viele aufwendige Filme mit großem Budget am selben Tag (wie etwa *Memorial Day* und *Independence Day*) in die Kinos kommen, antwortete er:

> Vermessenheit. Vermessenheit. Wenn man nur über sein eigenes Unternehmen nachdenkt, sagt man sich: »Ich hab eine gute Story-Abteilung, ich hab eine gute Marketing-Abteilung, wir bringen das jetzt raus.« Und man macht sich nicht klar, dass alle anderen genauso denken. An einem bestimmten Wochenende in einem Jahr laufen fünf Filme an, und da sind bestimmt nicht genügend Leute unterwegs, um die Kinos vollzukriegen.

Die freimütige Antwort verweist auf Vermessenheit, aber sie verrät keinen Überlegenheitsdünkel gegenüber konkurrierenden Studios. Die Wettbewerber gehen einfach nicht mit in die Entscheidung ein, bei der einmal mehr eine schwierige Frage durch eine leichte ersetzt wurde. Die zu beantwortende Frage lautet: Wie viele Menschen werden sich unseren Film ansehen, unter Berücksichtigung dessen, was unsere Konkurrenten tun? Die Frage, die sich das Management des Studios stattdessen stellte, ist einfacher und bezieht sich auf Informationen, die ihm unmittelbar zugänglich ist: Haben wir einen guten Film und eine gute Vermarktungsorganisation? Die vertrauten System-1-Prozesse der eingeschränkten Informationsverarbeitung (WYSIATI) und der Ersetzung führen sowohl zur Vernachlässigung der Konkurrenz als auch zum Above-Average-Effekt. Die Folge der Konkurrenzvernachlässigung ist ein überhöhter Markteintritt: Mehr Wettbewerber treten in den Markt ein, als sich langfristig profitabel darin behaupten können, sodass sie im Schnitt einen Verlust einfahren. [15] Dieses Ergebnis ist für den typischen neuen Wettbewerber am Markt enttäuschend, aber der gesamtwirtschaftliche Effekt könnte durchaus positiv sein. Tatsächlich nennen Giovanni Dosi und Dan Lovallo wagemutige Firmen, die scheitern, aber besser qualifizierte Konkurrenten auf neue Märkte hinweisen, »optimistische Märtyrer« – sie sind gut für die Wirtschaft insgesamt, aber schlecht für ihre Investoren.

## Selbstüberschätzung

Mehrere Jahre lang führten Professoren der Duke University eine Erhebung durch, bei der die Finanzvorstände von Großunternehmen die durchschnittliche Rendite eines Aktienindex von Standard & Poor's im folgenden Jahr vorhersagen sollten. Die Duke-Wissenschaftler sammelten 11 600 derartige Vorhersagen ein und überprüften ihre Genauigkeit. Das Ergebnis war eindeutig: Finanzvorstände von Großunternehmen hatten keinen blassen Schimmer davon, wie sich der Aktienmarkt auf kurze Sicht entwickeln würde; die Korrelation zwischen ihren Schätzungen und dem tatsächlichen Wert lag knapp unter null! Wenn sie sinkende Kurse vorhersagten, stiegen diese tatsächlich tendenziell eher an. Diese Ergebnisse sind nicht überraschend. Die eigentlich schlechte Nachricht ist, dass die Finanzvorstände anscheinend nicht wussten, dass ihre Vorhersagen wertlos sind.

Zusätzlich zu ihren Schätzungen über die Wertentwicklung des S&P-Aktienindex gaben die Befragten zwei weitere Schätzungen ab: einen Wert, der ihrer Einschätzung nach mit 90-prozentiger Sicherheit zu hoch war, und einen anderen, der mit ebensolcher Sicherheit zu niedrig war. Die Spanne zwischen den beiden Werten wird »80-Prozent-Konfidenzintervall« genannt, und Ergebnisse, die außerhalb dieses Intervalls liegen, werden als »Überraschungen« bezeichnet. Eine Person, die bei mehreren Gelegenheiten Konfidenzintervalle festlegt, geht davon aus, dass etwa 20 Prozent der Ergebnisse Überraschungen sind. Wie es bei solchen Übungen häufig geschieht, gab es viel zu viele Überraschungen; ihre Häufigkeit betrug 67 Prozent, über dreimal mehr als erwartet. Dies zeigt, dass Finanzvorstände ihre Fähigkeit, die Marktentwicklung vorherzusagen, stark überschätzten. Selbstüberschätzung ist eine weitere Manifestation der WYSIATI-Regel: Wenn wir eine Größe schätzen, stützen wir uns auf Informationen, die uns spontan einfallen, und konstruieren eine kohärente Geschichte, in der der Schätzwert sinnvoll erscheint. Informationen zu berücksichtigen, die einem nicht einfallen – vielleicht weil man sie nie wusste –, ist unmöglich.

Die Autoren berechneten die Konfidenzintervalle, die die Häufigkeit von überraschenden Werten auf 20 Prozent reduziert hätten. Die Ergebnisse

waren verblüffend. Um die Häufigkeit von überraschenden Werten auf dem gewünschten Niveau zu halten, hätten die Finanzvorstände Jahr für Jahr sagen sollen: »Die Wertentwicklung des S&P-Aktienindex wird im nächsten Jahr mit 80-prozentiger Wahrscheinlichkeit zwischen minus 10 und plus 30 Prozent liegen.« Das Konfidenzintervall, das das Wissen (genauer gesagt: das Unwissen) angemessen widerspiegelt, ist viermal breiter als die Intervalle, die sie nannten.

Hier kommt die Sozialpsychologie ins Spiel, weil die Antwort, die ein ehrlicher Finanzvorstand geben würde, schlicht lächerlich ist. Ein Finanzmanager, der seinen Kollegen mitteilen würde, die Rendite des S&P-Aktienindex liege mit hoher Wahrscheinlichkeit zwischen minus 10 und plus 30 Prozent, dürfte nur schallendes Gelächter ernten. Das breite Konfidenzintervall ist ein Eingeständnis der Unkenntnis, das für jemanden, der dafür bezahlt wird, sich in finanziellen Angelegenheiten gut auszukennen, gesellschaftlich nicht annehmbar ist. Selbst wenn diese Manager wüssten, wie wenig sie wissen, würden die Führungskräfte dafür bestraft, es zuzugeben. Präsident Truman wünschte sich bekanntermaßen einmal einen »einseitigen Ökonomen«, der einen klaren Standpunkt vertritt; er hatte die Nase voll von Ökonomen, die immer wieder sagten: »Andererseits …«

Organisationen, die sich selbst überschätzende Experten beim Wort nehmen, kann dies teuer zu stehen kommen. Die Studie über Finanzvorstände zeigte, dass diejenigen, die die Entwicklung des S&P-Index am optimistischsten beurteilten, zugleich die Aussichten ihres eigenen Unternehmens allzu optimistisch betrachteten, mit der Folge, dass es höhere Risiken einging als andere. Wie Nassim Taleb darlegte, veranlasst die unzulängliche Berücksichtigung der Unsicherheit des Umfeldes ökonomische Akteure zwangsläufig dazu, Risiken einzugehen, die sie vermeiden sollten. Doch Optimismus wird hoch geschätzt, sowohl gesellschaftlich als auch am Markt; Menschen und Firmen belohnen diejenigen stärker, die gefährlich irreführende Informationen bereitstellen, als diejenigen, die die Wahrheit sagen. Eine der Lektionen der Finanzkrise, die zur Großen Rezession führte, besteht darin, dass es Zeiten gibt, zu denen die Konkurrenz unter Experten und zwischen Organisationen starke Kräfte hervorbringt, die eine kollektive Blindheit gegenüber Risiken und Unsicherheit fördert.

Der soziale und wirtschaftliche Druck, der Selbstüberschätzung begünstigt, ist nicht auf Finanzprognosen beschränkt. Andere Fachkräfte müssen mit der Tatsache zurechtkommen, dass von einem Experten, der diesen Namen verdient, erwartet wird, ein hohes Maß an Selbstvertrauen zu zeigen. Philip Tetlock beobachtete, dass die Experten mit der stärksten Selbstüberschätzung am ehesten eingeladen wurden, in Nachrichtensendungen zu zeigen, was sie draufhaben. Auch in der Medizin scheint Selbstüberschätzung weit verbreitet zu sein. Bei einer Studie über Patienten, die auf der Intensivstation gestorben waren, wurden die Obduktionsbefunde mit der Diagnose verglichen, die Ärzte gestellt hatten, als die Patienten noch am Leben gewesen waren. Die Ärzte gaben auch das Ausmaß ihrer Urteilssicherheit an. Das Ergebnis: »Kliniker, die sich der Diagnose vor dem Tod ›vollkommen sicher‹ waren, irrten sich in 40 Prozent der Fälle.«[16] Auch hier wird die Selbstüberschätzung der Experten von ihren Klienten gefördert: »Im Allgemeinen gilt es als Schwäche und als ein Zeichen mangelhafter Qualifikation, wenn Kliniker unsicher erscheinen. Selbstsicherheit zählt mehr als Unsicherheit, und es gibt eine weitverbreitete Zensur gegen die Offenlegung von Unsicherheit gegenüber Patienten.«[17] Experten, die das ganze Ausmaß ihres Unwissens eingestehen, können damit rechnen, dass sie von selbstbewussteren Konkurrenten ersetzt werden, die eher in der Lage sind, das Vertrauen von Klienten zu gewinnen. Eine unvoreingenommene Einschätzung der Unsicherheit ist ein Eckpfeiler rationalen Entscheidungsverhaltens – aber es ist nicht das, was Menschen und Organisationen wollen. Extreme Unsicherheit ist unter gefährlichen Umständen lähmend, und das Eingeständnis, dass man lediglich Vermutungen anstellt, ist besonders unannehmbar, wenn viel auf dem Spiel steht. Da hält man sich oftmals lieber an vorgebliches Wissen.

Wenn emotionale, kognitive und soziale Faktoren, die überzogenen Optimismus unterstützen, zusammenkommen, sind sie eine berauschende Mischung, die Menschen manchmal dazu veranlasst, Risiken einzugehen, die sie meiden würden, wenn sie über die Erfolgschancen Bescheid wüssten. Dies ist kein Beweis dafür, dass risikofreudige Personen in der Wirtschaft eine ungewöhnlich große Lust auf Glücksspiele mit hohen Einsätzen hätten; sie sind sich lediglich der Risiken weniger deutlich bewusst, als es ängstlichere Menschen sind. Dan Lovallo und ich prägten

die Formel »kühne Vorhersagen und zaghafte Entscheidungen«, um den Hintergrund der Risikobereitschaft zu beschreiben. [18]

Die Auswirkungen von hohem Optimismus auf die Entscheidungsfindung sind bestenfalls ein zweifelhafter Segen, aber Optimismus leistet zweifellos einen positiven Beitrag zu einer guten Umsetzung. Der Hauptvorteil von Optimismus besteht in der Widerstandsfähigkeit gegenüber Rückschlägen. Laut Martin Seligman, dem Begründer der Positiven Psychologie, fördert ein »optimistischer Erklärungsstil« die seelische Widerstandskraft, indem er das eigene Selbstbild verteidigt. Der optimistische Stil besteht im Wesentlichen darin, sich Erfolge als eigenes Verdienst anzurechnen, während man sich für Misserfolge nicht tadelt. Dieser Stil kann zumindest bis zu einem gewissen Grad unterrichtet werden, und Seligman hat die Auswirkungen entsprechender Schulungen auf verschiedene Berufstätigkeiten dokumentiert, die mit einer hohen Misserfolgsrate einhergehen, wie etwa die Kaltakquise potenzieller Kunden durch Versicherungsvertreter (eine gängige Praxis vor dem Aufkommen des Internets). Wenn einem gerade von einer wütenden Hausfrau die Tür vor der Nase zugeknallt wurde, ist der Gedanke, dass »sie eine schreckliche Frau ist« viel angenehmer als der Gedanke, ein unfähiger Vertreter zu sein. Ich war schon immer der Überzeugung, dass wissenschaftliche Forschung ein weiteres Feld darstellt, in dem eine Form von Optimismus ausschlaggebend für den Erfolg ist: Bis jetzt bin ich noch keinem erfolgreichen Wissenschaftler begegnet, dem die Fähigkeit fehlt, die Bedeutung seiner Arbeit hochzuspielen. Und ich glaube, dass jemand, der nicht in beinahe wahnhafter Weise von sich überzeugt ist, von den wiederholten Erfahrungen vieler kleiner Misserfolge und seltener Triumphe, die das Schicksal der meisten Forscher sind, zermürbt wird.

## Wie die Prä-mortem-Methode helfen kann

Lässt sich übermäßiger Optimismus durch Training überwinden? Da bin ich nicht optimistisch. Es gab zahlreiche Versuche, Menschen beizubringen, Konfidenzintervalle anzugeben, die der Ungenauigkeit ihrer Urteile Rechnung tragen – doch nur wenige dieser Versuche berichteten von bescheidenen Erfolgen. Ein oft zitiertes Beispiel betrifft Geologen von Royal Dutch Shell, die bei ihren Beurteilungen potenzieller Bohrstellen größere Unsicherheiten einräumten, nachdem ihnen zahlreiche Fälle aus der Vergangenheit vorgestellt worden waren, bei denen das Ergebnis bekannt war.[19] In anderen Situationen wurde die Selbstüberschätzung dadurch abgeschwächt (aber nicht beseitigt), dass man Richter dazu ermunterte, konkurrierende Hypothesen in Betracht zu ziehen. Allerdings ist Selbstüberschätzung eine direkte Folge von System-1-Merkmalen, die abgeschwächt, aber nicht überwunden werden können. Das Haupthindernis besteht darin, dass der subjektive Grad des Überzeugtseins von der Kohärenz der Geschichte bestimmt wird, die man konstruiert hat, nicht von der Güte und Menge der Informationen, die sie stützen.

Organisationen mögen Optimismus besser zähmen können als Einzelpersonen. Die beste Idee dafür stammt von Gary Klein, meinem »gegnerischen Forschungspartner«, der im Allgemeinen die intuitive Entscheidungsfindung gegen Kritik, diese liefere verzerrte Ergebnisse, verteidigt und die Anwendung von Algorithmen im Regelfall ablehnt. Er bezeichnet seinen Ansatz als »Prä-mortem-Methode«. Die Vorgehensweise ist einfach: Wenn die Organisation kurz davorsteht, eine wichtige Entscheidung zu treffen, aber noch keinen förmlichen Beschluss gefasst hat, schlägt Klein vor, eine Gruppe von Personen, die bestens mit der Entscheidung vertraut sind, zu einer kurzen Sitzung zusammenzurufen. Die Sitzung beginnt mit einer kurzen Ansprache: »Stellen Sie sich vor, wir befinden uns ein Jahr in der Zukunft. Wir haben den Plan in seiner jetzigen Fassung umgesetzt. Das Ergebnis war eine Katastrophe. Nehmen Sie sich bitte fünf bis zehn Minuten Zeit, um eine kurze Geschichte dieser Katastrophe zu schreiben.«

Gary Kleins Idee der Prä-mortem-Analyse stößt für gewöhnlich auf sofortige Begeisterung. Nachdem ich sie bei einer Sitzung in Davos

beiläufig beschrieb, murmelte jemand hinter mir: »Schon allein deshalb hat es sich gelohnt, nach Davos zu kommen!« (Später erkannte ich, dass der Sprecher der Vorstandschef eines internationalen Konzerns war.) Die Prä-mortem-Analyse hat zwei entscheidende Vorteile: Sie überwindet das Gruppendenken, das sich auf viele Teams auswirkt, sobald eine Entscheidung gefallen zu sein scheint, und sie lenkt die Fantasie sachkundiger Personen in eine dringend benötigte Richtung.

In dem Maße, wie sich ein Team auf eine Entscheidung einigt – und insbesondere dann, wenn der Teamleiter seine Meinung kundtut –, werden öffentlich geäußerte Zweifel an der Vorteilhaftigkeit der geplanten Maßnahme allmählich unterdrückt und schließlich sogar als Beleg für die fehlende Loyalität gegenüber dem Team und seinen Anführern behandelt. Die Unterdrückung von Zweifeln trägt zur Selbstüberschätzung in einer Gruppe bei, in der nur Befürworter der Entscheidung ein Mitspracherecht haben. Der wichtigste Vorzug der Prä-mortem-Methode besteht darin, dass sie Zweifel zulässt. Außerdem ermuntert sie auch Befürworter der Entscheidung, nach möglichen Gefahren zu suchen, die sie bis dahin nicht in Betracht gezogen haben. Die Prä-mortem-Methode ist kein Allheilmittel, und sie bietet keinen vollständigen Schutz vor hässlichen Überraschungen, aber immerhin begrenzt sie das Schadenspotenzial von Plänen, die den Verzerrungen der Informationsverfügbarkeit (WYSIATI-Regel) und eines unkritischen Optimismus unterliegen.

## Zum Thema »Optimismus«

»Sie haben eine Kontrollillusion. Sie unterschätzen die Hindernisse ganz erheblich.«

»Sie scheinen an einem akuten Fall von Konkurrenzvernachlässigung zu leiden.«

»Dies ist ein Fall von Selbstüberschätzung. Sie scheinen zu glauben, dass sie mehr wissen, als sie es tatsächlich tun.«

»Wir sollten eine Prä-mortem-Sitzung durchführen. Vielleicht fällt jemandem eine Bedrohung ein, die wir vernachlässigt haben.«

# TEIL IV

## Entscheidungen

# *25. Irrtümer*

Eines Tages Anfang der 1970er-Jahre überreichte mir Amos die Kopie eines Aufsatzes des Schweizer Ökonomen Bruno Frey, in dem die psychologischen Annahmen der volkswirtschaftlichen Theorie diskutiert wurden. Ich habe noch genau die Farbe des Umschlags vor Augen: dunkelrot. Bruno Frey erinnert sich kaum mehr an diesen Aufsatz, aber ich kenne seinen ersten Satz noch immer auswendig: »Der Agent der volkswirtschaftlichen Theorie ist rational, egoistisch, und seine Präferenzen verändern sich nicht.«

Ich war erstaunt. Meine Kollegen vom wirtschaftswissenschaftlichen Fachbereich arbeiteten in dem Gebäude nebenan, aber ich hatte nicht geahnt, wie groß der Unterschied zwischen unseren intellektuellen Welten ist. Für einen Psychologen versteht es sich von selbst, dass Menschen weder vollkommen rational noch völlig egoistisch sind und dass ihre Präferenzen alles andere als stabil sind. Unsere beiden Disziplinen schienen verschiedene Spezies zu erforschen, die der Verhaltensökonom Richard Thaler später *Econs* (Homo oeconomicus) und *Humans* (Normalsterbliche) nannte.

Anders als *Econs* haben die *Humans*, die von Psychologen erforscht werden, ein System 1. Ihr Weltbild wird durch die Informationen eingeschränkt, die zum jeweiligen Zeitpunkt verfügbar sind (WYSIATI), und daher können sie nicht so konsistent und logisch wie *Econs* sein. Sie sind manchmal großzügig und oftmals bereit, etwas für die Gruppe zu tun, mit der sie verbunden sind. Und sie haben häufig keine genaue Vorstellung davon, was sie nächstes Jahr oder auch morgen mögen werden. Hier bot sich also die Gelegenheit zu einem interessanten interdisziplinären Gespräch. Ich ahnte nicht, dass meine weitere wissenschaftliche Laufbahn maßgeblich von diesem Gespräch beeinflusst werden sollte.

Bald nachdem mir Amos Freys Aufsatz gezeigt hatte, schlug er vor, wir sollten das Studium der Entscheidungsfindung zu unserem nächsten gemeinsamen Projekt machen. Ich wusste so gut wie nichts über dieses Thema, aber Amos war ein Experte und ein Star auf diesem Gebiet, und er sagte, er würde mich coachen. Noch als Doktorand war er einer der

Mitverfasser des Lehrbuchs *Mathematical Psychology*, und er empfahl mir die Lektüre einiger Kapitel, die er für eine gute Einführung in das Thema hielt.[1]

Schon bald erfuhr ich, dass wir die Einstellungen von Menschen zu riskanten Optionen erforschen und nach einer Antwort auf eine konkrete Frage suchen würden: Nach welchen Regeln wählen Menschen zwischen verschiedenen einfachen Lotterien (Glücksspielen) und zwischen Lotterien und sicheren Optionen?

Einfache Lotterien (wie etwa »eine 40-prozentige Chance, 300 Dollar zu gewinnen«) sind für Forscher auf dem Gebiet der Entscheidungsfindung das, was die Fruchtfliege für Genetiker ist. Die Entscheidungen zwischen solchen Glücksspielen liefern ein einfaches Modell, das in wichtigen Merkmalen mit jenen komplexeren Entscheidungen übereinstimmt, die Forscher verstehen wollen. In Lotterien spiegelt sich die Tatsache wider, dass die Folgen von Entscheidungen niemals sicher vorhersehbar sind. Selbst scheinbar sichere Ergebnisse sind ungewiss: Wenn man den Kaufvertrag für eine Wohnung unterzeichnet, kennt man nicht den Preis, zu dem man sie später vielleicht verkaufen muss, und man weiß auch nicht, dass der Sohn des Nachbarn bald damit anfangen wird, Tuba zu üben. Jede wichtige Entscheidung, die wir im Leben treffen, geht mit einer bestimmten Ungewissheit einher – aus diesem Grund hoffen Wissenschaftler, die das menschliche Entscheidungsverhalten erforschen, dass einige der Erkenntnisse, die sich in Modellsituationen gewinnen lassen, auf interessantere Alltagsprobleme anwendbar sind. Aber natürlich besteht für Entscheidungstheoretiker die hauptsächliche Motivation zur Erforschung einfacher Glücksspiele darin, dass andere Entscheidungstheoretiker das Gleiche tun.

Die Disziplin verfügte über eine Theorie, die Erwartungsnutzentheorie, die die Grundlage des Modells vom rationalen Agenten bildet und die bis heute die wichtigste Theorie in den Sozialwissenschaften ist. Diese Theorie wurde nicht als psychologisches Modell entwickelt; sie sollte vielmehr die Logik von Wahlhandlungen auf der Basis elementarer Rationalitätsregeln (Axiome) beschreiben. Betrachten Sie dieses Beispiel:

> Wenn Sie einen Apfel einer Banane vorziehen,
> dann

ziehen Sie auch eine 10-prozentige Chance, einen Apfel zu gewinnen,
einer 10-prozentigen Chance vor, eine Banane zu gewinnen.

Apfel und Banane stehen für beliebige Wahlobjekte (einschließlich Lotterien), und die 10-prozentige Chance steht für jede beliebige Wahrscheinlichkeit. Der Mathematiker John von Neumann, eine der Geistesgrößen des 20. Jahrhunderts, und der Ökonom Oskar Morgenstern haben ihre Theorie der rationalen Wahl zwischen Lotterien aus einigen wenigen Axiomen abgeleitet. Die Ökonomen haben die Theorie des erwarteten Nutzens in einer doppelten Funktion übernommen: als eine Logik, die vorschreibt, wie Entscheidungen getroffen werden sollten, und als Beschreibung des Entscheidungsverhaltens der *Econs*. Doch Amos und ich waren Psychologen, und wir wollten verstehen, wie *Humans* riskante Entscheidungen treffen, ohne irgendwelche Annahmen über ihre Rationalität zu machen.

Wir behielten unsere Routine bei, tagtäglich viele Stunden mit Gesprächen zu verbringen, manchmal in unseren Büros, manchmal in Restaurants, oftmals auf langen Spaziergängen durch ruhige Straßen des schönen Jerusalems. Wie schon bei der Erforschung des Urteilsverhaltens haben wir auch hier unsere eigenen intuitiven Präferenzen gründlich analysiert. Wir verbrachten unsere Zeit damit, einfache Entscheidungsprobleme zu erfinden und uns selbst zu fragen, wie wir uns entscheiden würden. Zum Beispiel:

Was würden Sie vorziehen?

a. Werfen Sie eine Münze. Bei Kopf gewinnen Sie 100 Dollar, bei Zahl gewinnen Sie nichts.
b. Sie bekommen sicher 46 Dollar.

Wir versuchten nicht, die rationalste oder vorteilhafteste Wahl herauszufinden; wir wollten die intuitive Entscheidung ermitteln, jene, die unmittelbar verlockend ist. Wir wählten fast immer die gleiche Option aus.

In diesem Beispiel hätten wir uns beide für die sichere Alternative entschieden, und Sie würden vermutlich das Gleiche tun. Wenn wir uns mit großer innerer Überzeugung auf eine Option verständigten, glaubten wir – fast immer zutreffend, wie sich herausstellte –, dass die meisten Menschen unsere Präferenz teilen würden, und wir machten weiter, als hätten wir stichhaltige Beweise dafür. Natürlich wussten wir, dass wir unser Bauchgefühl später überprüfen müssten, aber dadurch, dass wir sowohl Experimentatoren als auch Probanden waren, kamen wir zügig voran.

Fünf Jahre nachdem wir mit der Erforschung von Lotterien begonnen hatten, beendeten wir schließlich einen Aufsatz mit dem Titel »Prospect Theory: An Analysis of Decision under Risk« (»Neue Erwartungstheorie: Eine Analyse risikoträchtiger Entscheidungen«). Unsere Theorie orientierte sich am Vorbild der Nutzentheorie, von der sie jedoch in grundlegender Weise abwich. Am wichtigsten aber ist, dass unser Modell rein deskriptiv war und dass es systematische Verstöße gegen die Axiome rationalen Handelns bei Wahlen zwischen Lotterien dokumentieren und erklären sollte. Wir reichten unseren Aufsatz bei *Econometrica* ein, einem Wissenschaftsjournal, das bedeutende theoretische Beiträge zur Ökonomie und Entscheidungstheorie veröffentlicht. Die Wahl des Publikationsorgans erwies sich als wichtig; hätten wir den Aufsatz in einer psychologischen Fachzeitschrift veröffentlicht, hätte er vermutlich keinen großen Einfluss auf die Wirtschaftswissenschaften gehabt. Doch unsere Entscheidung war nicht von dem Wunsch geleitet, Einfluss auf die Wirtschaftswissenschaften zu nehmen; *Econometrica* war einfach die Zeitschrift, in der die besten Aufsätze zur Entscheidungsforschung veröffentlicht worden waren, und wir wollten in dieser Gesellschaft sein. Bei dieser Wahl wie bei vielen anderen hatten wir Glück. Die Neue Erwartungstheorie *(Prospect Theory)* erwies sich als unsere bedeutendste wissenschaftliche Arbeit überhaupt, und unser Artikel gehört zu den meistzitierten Aufsätzen in den Sozialwissenschaften. Zwei Jahre später veröffentlichten wir in *Science* einen Aufsatz über Framing-Effekte: die weitreichenden Präferenzänderungen, die manchmal durch geringfügige Abwandlungen der Formulierung eines Entscheidungsproblems verursacht werden.

In den ersten fünf Jahren, in denen wir erforschten, wie Menschen Entscheidungen treffen, förderten wir ein Dutzend Fakten über Entscheidungen zwischen riskanten Optionen zutage. Mehrere dieser Fakten standen in eklatantem Widerspruch zur Erwartungsnutzentheorie.

Einige davon waren bereits beobachtet worden, andere waren neu. Dann konstruierten wir eine Theorie, die die Erwartungsnutzentheorie gerade so weit modifizierte, dass sie unsere Sammlung von Beobachtungen erklärte. Das war die Neue Erwartungstheorie.

Unsere Herangehensweise an das Problem war von dem Geist eines Teilgebiets der Psychologie geprägt, der sogenannten Psychophysik, die von dem deutschen Psychologen und Mystiker Gustav Theodor Fechner (1801–1887) begründet und als solche benannt wurde. Fechner war wie besessen von der Beziehung zwischen Geist und Materie. Auf der einen Seite gibt es eine veränderliche physikalische Größe, wie etwa die Energie einer Lichtquelle, die Frequenz eines Tons oder einen Geldbetrag. Andererseits gibt es das subjektive Erleben von Helligkeit, Tonhöhe oder Wert. Schwankungen der physikalischen Größe verursachen auf rätselhafte Weise Schwankungen in der Intensität oder Qualität des subjektiven Erlebens. Fechners Vorhaben bestand darin, die psychophysischen Gesetze zu finden, die einen Zusammenhang herstellen zwischen der subjektiven Größe im Bewusstsein des Beobachters und der objektiven Größe in der materiellen Welt. Er behauptete, bei vielen Dimensionen handele es sich um eine logarithmische Funktion – was schlichtweg bedeutet, dass eine Erhöhung der Reizintensität um einen bestimmten Faktor (zum Beispiel um 1,5 oder 10) immer den gleichen Zuwachs auf der psychologischen Skala ergibt. Wenn die Steigerung der Energie des Schalls von zehn auf hundert Einheiten physikalischer Energie die psychologische Intensität um vier Einheiten steigert, dann geht eine weitere Erhöhung der Reizintensität von hundert auf tausend ebenfalls mit einer Steigerung der psychologischen Intensität um vier Einheiten einher.

## Bernoullis Irrtum

Wie Fechner ganz genau wusste, war er nicht der Erste auf der Suche nach einer Funktion, die den Zusammenhang zwischen der psychologischen Intensität und der physikalischen Stärke des Reizes beschrieb. Im Jahr 1738 hatte der Schweizer Naturwissenschaftler Daniel Bernoulli Fechners Überlegung vorweggenommen und sie auf den Zusammenhang zwischen dem psychologischen Wert oder der Begehrtheit von Geld (heute »Nutzen« genannt) und dem tatsächlichen Geldbetrag angewandt. Er behauptete, ein Geschenk in Höhe von 10 Dukaten habe für jemanden, der bereits 100 Dukaten besitze, den gleichen Nutzen wie ein Geschenk von 20 Dukaten für jemanden, dessen gegenwärtiges Vermögen 200 Dukaten betrage. Natürlich hatte Bernoulli recht: Normalerweise geben wir Einkommensänderungen in Prozent an, so sagen wir zum Beispiel: »Sie bekam eine 30-prozentige Gehaltserhöhung.« Dahinter steht die Annahme, dass eine 30-prozentige Steigerung bei Reichen und bei Armen eine ganz ähnliche psychologische Wirkung hervorrufen wird, was bei einer Erhöhung um 100 Dollar nicht der Fall ist.[2] Wie beim Fechnerschen Gesetz ist die psychologische Reaktion auf eine Vermögensänderung umgekehrt proportional zum anfänglichen Vermögenszustand, was zu der Schlussfolgerung führt, dass der Nutzen eine logarithmische Funktion des Vermögens ist. Wenn diese Funktion zutreffend ist, dann trennt die gleiche psychologische Distanz 100 000 von 1 Million Dollar und 10 Millionen von 100 Millionen Dollar.[3] Ausgehend von seinen psychologischen Erkenntnissen über den Nutzen von Reichtum, schlug Bernoulli einen völlig neuen Ansatz zur Bewertung von Lotterien vor, ein wichtiges Thema für die Mathematiker seiner Zeit. Vor Bernoulli waren Mathematiker davon ausgegangen, dass Lotterien nach ihrem Erwartungswert beurteilt werden müssten: einem gewichteten Mittelwert der möglichen Ergebnisse, wobei jedes Ergebnis mit seiner Wahrscheinlichkeit gewichtet wird.

> So beträgt zum Beispiel der Erwartungswert einer 80-prozentigen Chance, 100 Dollar zu gewinnen, und einer 20-

prozentigen Chance, 10 Dollar zu gewinnen, 82 Dollar (0,8 × 100 + 0,2 × 10).

Stellen Sie sich jetzt selbst folgende Frage: Was würden Sie lieber als Geschenk bekommen: diese Lotterie oder 80 Dollar sicher? Fast alle bevorzugen die sichere Option. Wenn Menschen unsichere Aussichten nach ihrem Erwartungswert beurteilen würden, dann würden sie die Lotterie vorziehen, weil 82 Dollar mehr sind als 80 Dollar. Bernoulli wies darauf hin, dass Menschen Lotterien tatsächlich nicht in dieser Weise beurteilen.

Bernoulli beobachtete, dass die meisten Menschen risikoscheu sind (Risiko = die Wahrscheinlichkeit, den niedrigsten möglichen Geldbetrag zu erhalten), und wenn sie die Wahl zwischen einem Glücksspiel und einem Betrag haben, der dem Erwartungswert des Glücksspiels entspricht, dann entscheiden sie sich für die sichere Option. Tatsächlich wird sich ein risikoscheuer Entscheider sogar für eine sichere Option entscheiden, wenn deren Wert unter dem Erwartungswert der Lotterie liegt, und er wird sogar eine Prämie bezahlen, um die Unsicherheit zu vermeiden. Einhundert Jahre vor Fechner erfand Bernoulli die Psychophysik, um diese Risikoaversion zu erklären. Seine Idee war einfach: Menschen lassen sich bei ihren Wahlhandlungen nicht von den monetären Werten, sondern von den psychologischen Werten der Ergebnisse, ihrem Nutzen, leiten. Daher ist der psychologische Wert einer Lotterie nicht der gewichtete Mittelwert ihrer möglichen monetären Ergebnisse, sondern der Mittelwert aller Nutzwerte dieser mit ihren jeweiligen Wahrscheinlichkeiten gewichteten Ergebnisse.

| Vermögen (in Mio.) | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 |
|---|---|---|---|---|---|---|---|---|---|---|
| Nutzeneinheiten | 10 | 30 | 48 | 60 | 70 | 78 | 84 | 90 | 96 | 100 |

**Tabelle 3**

In Tabelle 3 ist eine Variante der von Bernoulli berechneten Nutzenfunktion zu sehen; sie zeigt den Nutzen verschiedener Vermögensbeträge, von 1 Million bis 10 Millionen. Sie sehen, dass die Erhöhung eines Vermögens von 1 Million um 1 Million einen Zuwachs von zwanzig Nutzenpunkten ergibt, während die Aufstockung eines Vermögens von 9 Millionen um 1 Million nur vier Punkte bringt. Bernoulli behauptete, der rückläufige

Grenznutzen des Vermögens (im modernen ökonomischen Jargon) erkläre die Risikoaversion – die weitverbreitete Präferenz für ein sicheres Ergebnis gegenüber einer günstigen Lotterie mit einem gleich hohen oder geringfügig höheren Erwartungswert. Betrachten wir diese Wahl:

| | |
|---|---|
| Die gleiche Chance, 1 Million oder | Nutzen: (10 + 84)/2 = 47 |
| 7 Millionen zu gewinnen. | |
| Oder: | |
| 4 Millionen sicher haben. | Nutzen: 60 |

Die Erwartungswerte der Lotterie und der »sicheren Sache« sind in Dukaten gleich groß (4 Millionen), aber der psychologische Nutzen der beiden Optionen ist unterschiedlich, wegen des abnehmenden Nutzwerts von Vermögen: Der Nutzenzuwachs von 1 Million auf 4 Millionen beträgt fünfzig Einheiten, aber ein gleich großer Vermögenszuwachs von 4 auf 7 Millionen steigert den Nutzen des Vermögens nur um 24 Einheiten. Der Nutzen der Lotterie ist 94/2 = 47 (der Nutzen der beiden Ergebnisse, jeweils gewichtet mit seiner Wahrscheinlichkeit von 50 Prozent). Der Nutzen von 4 Millionen ist sechzig. Weil das mehr als 47 ist, wird eine Person mit dieser Nutzenfunktion die sichere Option bevorzugen. Bernoulli erkannte, dass ein Entscheider bei rückläufigem Grenznutzen des Vermögens risikoscheu ist.

Bernoullis Aufsatz ist ein Meisterwerk an argumentativer Prägnanz. Er wandte sein neues Konzept des erwarteten Nutzens (den er »moralische Erwartung« nannte) an, um zu berechnen, wie viel ein Kaufmann in St. Petersburg zu zahlen bereit wäre, um eine Schiffsladung Gewürze aus Amsterdam zu versichern, wenn »er sich der Tatsache bewusst ist, dass zu dieser Jahreszeit von einhundert Schiffen, die von Amsterdam nach St. Petersburg segeln, in der Regel fünf verloren gehen«. Seine Nutzenfunktion erklärte, wieso arme Menschen Versicherungen abschließen und weshalb reichere Leute sie ihnen verkaufen. Wie Sie aus der Tabelle entnehmen können, erleidet jemand mit einem Vermögen von 10 Millionen, das um 1

Million schrumpft, einen Verlust von vier Nutzenpunkten (von 100 auf 96). Hingegen hat jemand, der von 3 Millionen 1 Million einbüßt, einen viel größeren Verlust zu verzeichnen, nämlich 18 Punkte (von 48 auf 30). Der ärmere Mann wird bereitwillig eine Prämie zahlen, um das Risiko auf den reicheren abzuwälzen, und genau darum geht es bei Versicherungen. Bernoulli wartete auch mit einer Lösung für das berühmte Sankt-Petersburg-Paradoxon auf, bei dem Menschen, denen eine Lotterie mit einem unendlichen Erwartungswert (in Dukaten) angeboten wird, nur ein paar Dukaten dafür auszugeben bereit sind.[4] Am eindrucksvollsten aber ist, dass seine Analyse der Risikoeinstellungen in Abhängigkeit von Vermögenspräferenzen den Test der Zeit bestanden hat: Fast 300 Jahre später ist sie in der ökonomischen Analyse noch immer gebräuchlich.

Die Langlebigkeit der Theorie ist umso bemerkenswerter, als sie mit gravierenden Fehlern behaftet ist. Die Irrtümer einer Theorie finden sich selten in ihren ausdrücklichen Behauptungen, vielmehr verstecken sie sich in dem, was sie ausblendet oder stillschweigend annimmt. Nehmen wir als Beispiel die folgenden Szenarien:

> Heute besitzen Jack und Jill jeweils ein Vermögen von 5 Millionen.
> Gestern hatte Jack 1 Million und Jill hatte 9 Millionen.
> Sind sie gleich zufrieden? (Haben sie den gleichen Nutzen?

Bernoullis Theorie nimmt an, dass der Nutzen ihres Vermögens das ist, was Menschen mehr oder weniger glücklich macht. Jack und Jill haben das gleiche Vermögen, und aus diesem Grund behauptet die Theorie, dass sie gleich zufrieden sein sollten. Aber man braucht kein Diplom in Psychologie, um zu wissen, dass Jack heute gute Laune hat, während Jill niedergeschlagen ist. Tatsächlich wissen wir, dass Jack viel zufriedener als Jill wäre, selbst wenn er heute nur 2 Millionen hätte, während sie 5 Millionen hat. Daher muss Bernoullis Theorie falsch sein.

Die Zufriedenheit, die Jack und Jill erleben, wird von der aktuellen *Veränderung* ihrer Vermögenslage im Verhältnis zu den verschiedenen Vermögenszuständen bestimmt, die ihre Referenzpunkte definieren (1

Million für Jack, 9 Millionen für Jill). Diese Referenzabhängigkeit ist bei Empfindung und Wahrnehmung allgegenwärtig. Das gleiche Geräusch wird als sehr laut oder recht leise erlebt, je nachdem, ob ihm ein Flüstern oder ein Brüllen vorausgeht. Um das subjektive Erleben der Lautstärke vorherzusagen, genügt es nicht, die absolute Energie des Schalls zu kennen; man muss auch die Referenzschallstärke kennen, mit der das Schallereignis automatisch verglichen wird. In ähnlicher Weise muss man den Hintergrund kennen, ehe man vorhersagen kann, ob ein grauer Fleck auf einer Seite dunkel oder hell erscheinen wird. Und so muss man auch die Bezugsgröße kennen, ehe man den Nutzen einer bestimmten Vermögenshöhe vorhersagen kann.

Ein weiteres Beispiel dafür, was Bernoullis Theorie entgeht, liefern Anthony und Betty:

> Anthonys gegenwärtiges Vermögen beträgt 1 Million.
> Bettys gegenwärtiges Vermögen beträgt 4 Millionen.

Beiden wird die Wahl zwischen einer Lotterie und einer sicheren Option angeboten:

> Die Lotterie: Mit jeweils gleich hoher Wahrscheinlichkeit am Ende 1 Million oder 4 Millionen besitzen.
>
> Oder:
>
> Am Ende 2 Millionen sicher besitzen.

Nach Bernoullis Theorie stehen Anthony und Betty vor der gleichen Wahl: Ihr erwartetes Vermögen beträgt 2,5 Millionen, wenn sie sich für das Glücksspiel entscheiden, und 2 Millionen, wenn sie die sichere Option vorziehen. Daher würde Bernoulli erwarten, dass Anthony und Betty die gleiche Wahl treffen, aber diese Vorhersage ist falsch. Die Theorie scheitert hier abermals, weil sie keine verschiedenen Referenzpunkte zulässt, von

denen aus Anthony und Betty ihre Optionen abwägen. Wenn Sie sich in Anthonys und Bettys Lage hineinversetzen, werden Sie schnell erkennen, dass das gegenwärtige Vermögen eine große Rolle spielt. Sie denken vermutlich folgendermaßen:

> Anthony (der gegenwärtig 1 Million besitzt): »Wenn ich mich für die sichere Option entscheide, verdoppele ich mein Vermögen mit Sicherheit. Das ist sehr attraktiv. Alternativ kann ich eine Wette eingehen, wobei ich mit je 50-prozentiger Wahrscheinlichkeit mein Vermögen vervierfachen oder nichts gewinnen werde.«

> Betty (die gegenwärtig 4 Millionen besitzt): »Wenn ich mich für die sichere Option entscheide, verliere ich mit Sicherheit die Hälfte meines Vermögens, was schrecklich ist. Alternativ kann ich eine Wette darauf eingehen, dass ich mit je 50-prozentiger Wahrscheinlichkeit drei Viertel meines Vermögens oder nichts verlieren werde.«

Sie ahnen schon, dass Anthony und Betty höchstwahrscheinlich unterschiedliche Entscheidungen treffen werden, weil die sichere Option, 2 Millionen zu besitzen, Anthony glücklich und Betty unglücklich macht. Beachten Sie auch, dass sich das *sichere* Ergebnis vom *schlechtesten* Ergebnis der Lotterie unterscheidet: Für Anthony ist es der Unterschied zwischen der Verdopplung seines Vermögens und einer gleichbleibenden Vermögenshöhe; für Betty ist es der Unterschied zwischen dem Verlust der Hälfte ihres Vermögens und dem Verlust von drei Vierteln ihres Vermögens. Betty wird mit viel höherer Wahrscheinlichkeit das Risiko eingehen, so wie es andere tun, wenn sie die Wahl zwischen sehr schlechten Optionen haben. So, wie ich ihre Geschichte geschildert habe, denken weder Anthony noch Betty in Kategorien von Vermögenszuständen: Anthony denkt an Gewinne, und Betty denkt an Verluste. Die von ihnen beurteilten psychologischen

Ergebnisse sind völlig unterschiedlich, obwohl die möglichen Vermögenszustände, mit denen sie konfrontiert sind, identisch sind.

Weil Bernoullis Modell das Konzept des Referenzpunktes fehlt, berücksichtigt die Erwartungsnutzentheorie nicht die offenkundige Tatsache, dass das Ergebnis, das für Anthony vorteilhaft ist, für Betty Nachteile bringt. Sein Modell könnte Anthonys Risikoaversion erklären, aber nicht Bettys risikofreudige Präferenz für die Lotterie, ein Verhalten, das sich oftmals bei Unternehmern und Generälen beobachten lässt, wenn sämtliche ihrer Optionen schlecht sind.

All dies scheint auf der Hand zu liegen, oder nicht? Man könnte sich ohne Weiteres vorstellen, dass Bernoulli selbst ähnliche Beispiele konstruiert und eine komplexere Theorie aufgestellt hätte, um ihnen Rechnung zu tragen; aus irgendeinem Grund tat er dies nicht. Man könnte sich auch vorstellen, dass ihm zeitgenössische Kollegen widersprochen oder spätere Wissenschaftler nach der Lektüre seines Aufsatzes Einwände erhoben haben, aber das geschah ebenfalls nicht.

Die Frage ist, wieso sich eine Konzeption des Nutzens von Ergebnissen, die mit so naheliegenden Gegenbeispielen angreifbar ist, so lange halten konnte. Ich kann es nur mit einer intellektuellen Schwäche vieler Wissenschaftler erklären, die ich häufig bei mir selbst beobachtet habe. Ich nenne sie »theorieinduzierte Blindheit«: Sobald man eine Theorie anerkannt hat und als intellektuelles Werkzeug benutzt, ist es überaus schwer, ihre Schwächen zu bemerken. Wenn man eine scheinbar nicht mit dem Modell in Einklang zu bringende Beobachtung macht, geht man davon aus, dass es eine sehr gute Erklärung dafür geben muss, die einem aus irgendeinem Grund entgeht. Im Zweifelsfall entscheidet man zugunsten der Theorie und vertraut der Gemeinschaft von Experten, die sie für richtig halten. Zweifellos gingen vielen Wissenschaftlern irgendwann einmal Fallbeispiele wie die von Anthony und Betty oder Jack und Jill durch den Kopf, und sie stellten beiläufig fest, dass diese Beispiele nicht mit der Nutzentheorie in Einklang zu bringen waren. Aber sie gingen ihren Bedenken nicht so weit nach, dass sie zu der Schlussfolgerung hätten kommen müssen: »Diese Theorie weist gravierende Mängel auf, weil sie die Tatsache ignoriert, dass der Nutzen von der Vorgeschichte des Vermögens einer Person abhängig ist, nicht nur von ihrem gegenwärtigen Vermögensstand.«[5] Der Psychologe Daniel Gilbert hat ganz richtig bemerkt, dass es mühsam ist, Dinge anzuzweifeln, und System 2 ist schnell erschöpft.

## Zum Thema »Bernoullis Irrtum«

»Er freute sich sehr, als er vor drei Jahren eine Prämie von 20 000 Dollar erhielt, aber seither ist sein Gehalt um 20 Prozent gestiegen, sodass er eine höhere Prämie braucht, um den gleichen Nutzen zu erhalten.«

»Beide Kandidaten sind mit dem Gehalt, das wir bieten, zufrieden, aber ihre Zufriedenheit wird nicht gleich groß sein, weil ihre Referenzpunkte unterschiedlich sind. Sie hat gegenwärtig ein viel höheres Gehalt.«

»Sie verklagt ihn auf Alimente. Sie würde sich gern gütlich mit ihm einigen, aber er prozessiert lieber. Das ist nicht weiter verwunderlich – sie kann nur gewinnen, sodass sie risikoscheu ist. Seine Optionen dagegen sind alle schlecht, sodass er lieber das Risiko eingeht.«

# *26. Die Neue Erwartungstheorie*

Dank einer glücklichen Kombination aus Können und Unwissenheit entdeckten Amos und ich die zentrale Schwachstelle in Bernoullis Theorie. Auf Amos' Empfehlung hin las ich ein Kapitel in seinem Buch. Es beschrieb Experimente, bei denen renommierte Wissenschaftler den Nutzen von Geld gemessen hatten, indem sie Menschen aufforderten, eine Auswahl unter Lotterien vorzunehmen, bei denen der Teilnehmer ein paar Penny gewinnen oder verlieren konnte. Die Experimentatoren maßen den Nutzwert von Vermögen, indem sie das Vermögen innerhalb einer Spanne von weniger als einem Dollar schwanken ließen. Dies warf Fragen auf. Ist die Annahme plausibel, dass Menschen die Chancen und Risiken von Glücksspielen anhand äußerst geringer Vermögensunterschiede beurteilen? Wie konnte man erwarten, etwas über die Psychophysik des Vermögens in Erfahrung zu bringen, indem man Reaktionen auf Gewinne und Verluste in Höhe von Pennybeträgen untersuchte? Neuere Entwicklungen in der Theorie der Psychophysik deuteten darauf hin, dass man direkte Fragen nach dem Vermögen, nicht aber nach Vermögensänderungen stellen sollte, wenn man den subjektiven Wert von Vermögenszuständen erforschen wollte.[1] Ich wusste nicht genug über die Nutzentheorie, um vor Respekt vor ihr geblendet zu sein, und ich war verdutzt.

Als Amos und ich uns am nächsten Tag trafen, berichtete ich ihm beiläufig, nicht auftrumpfend, von meinen Schwierigkeiten. Ich erwartete, dass er mich aufklären und mir erläutern würde, weshalb das Experiment, das mich verwirrt hatte, doch im Rahmen der Nutzentheorie erklärt werden konnte. Aber er tat nichts dergleichen – die Bedeutung der modernen Psychophysik leuchtete ihm sofort ein. Er erinnerte sich daran, dass der Wirtschaftswissenschaftler Harry Markowitz, der später für seine Beiträge zur Finanzwirtschaft mit dem Nobelpreis ausgezeichnet wurde, eine Theorie vorgeschlagen hatte, wonach Nutzwerte mit Vermögensänderungen und nicht mit Vermögenszuständen verbunden sind. Markowitz' Theorie war damals schon 25 Jahre alt und hatte nicht viel Beachtung gefunden, aber wir gelangten schon bald zu dem Schluss, dass diese Theorie in die richtige Richtung wies und dass die Theorie, die wir entwickeln wollten,

Ergebnisse als Gewinne und Verluste, nicht als Vermögenszustände definieren würde. Wahrnehmungspsychologische Kenntnisse und Unkenntnis der Entscheidungstheorie haben gemeinsam unsere Forschung ein großes Stück vorangebracht. Wir wussten schon bald, dass wir einen gravierenden Fall von theorieinduzierter Blindheit überwunden hatten, weil die Theorie, die wir verworfen hatten, jetzt nicht nur falsch, sondern absurd anmutete. Es erheiterte uns, zu erkennen, dass wir unser gegenwärtiges Vermögen nur innerhalb einer Ungenauigkeitsspanne von etlichen Zehntausend Dollar angeben konnten. Die Vorstellung, aus dem Vermögensnutzen Einstellungen zu geringfügigen Vermögensänderungen abzuleiten, erschien uns jetzt unhaltbar. Man weiß, dass man einen theoretischen Fortschritt erzielt hat, wenn man nicht mehr rekonstruieren kann, weshalb man das Offensichtliche so lange übersehen hat. Dennoch brauchten wir Jahre, um uns über die Folgen klar zu werden, die mit der Umorientierung auf Gewinne und Verluste verbunden waren.

In der Nutzentheorie wird der Nutzen eines Gewinns durch den Vergleich der Nutzen zweier Vermögenszustände beurteilt. So ist zum Beispiel der Nutzen zusätzlicher 500 Dollar bei einem Vermögen von 1 Million Dollar die Differenz zwischen dem Nutzen von 1 000 500 Dollar und dem Nutzen von 1 Million Dollar. Und wenn man den größeren Betrag besitzt, ist der negative Nutzen eines Verlusts von 500 Dollar wiederum die Differenz zwischen den Nutzen der beiden Vermögenszustände. In dieser Theorie unterscheiden sich die Nutzen von Gewinnen und Verlusten nur in ihren Vorzeichen (+ oder – ). Die Tatsache, dass der negative Nutzen eines Verlusts von 500 Dollar größer sein könnte als der Nutzen eines Gewinnes in gleicher Höhe, lässt sich in dieser Theorie nicht abbilden. Wie in einer Situation theorieinduzierter Blindheit zu vermuten, wurden Differenzen zwischen Gewinnen und Verlusten weder erwartet noch erforscht. Es wurde einfach angenommen, dass der Unterschied zwischen Gewinnen und Verlusten keine Rolle spiele, also war es auch nicht nötig, sich näher damit zu befassen.

Amos und ich erkannten nicht sofort, dass unsere Konzentration auf Vermögensveränderungen den Weg zur Erkundung eines neuen Themas ebnete. Wir interessierten uns hauptsächlich für Unterschiede zwischen Lotterien mit hohen oder niedrigen Gewinnwahrscheinlichkeiten. Eines Tages machte Amos einen beiläufigen Vorschlag: »Wie steht es mit Verlusten?« Wir fanden rasch heraus, dass unsere vertraute Risikoaversion

durch Risikobereitschaft ersetzt wurde, wenn wir einen Wechsel der Perspektive vollzogen. Betrachten wir diese beiden Probleme:

Problem 1: Was wählen Sie?
900 Dollar sicher erhalten oder eine 90-prozentige Chance, 1000 Dollar zu gewinnen.

Problem 2: Was wählen Sie?
Einen sicheren Verlust von 900 Dollar oder eine 90-prozentige Wahrscheinlichkeit, 1000 Dollar zu verlieren.

Vermutlich waren Sie bei Problem 1 risikoscheu, wie es die große Mehrzahl der Menschen ist. Der subjektive Wert eines Gewinns von 900 Dollar ist zweifellos größer als 90 Prozent des Wertes eines Gewinns von 1000 Dollar. Die risikoscheue Wahl bei diesem Problem hätte Bernoulli nicht überrascht.

Betrachten Sie nun Ihre Präferenz bei Problem 2. Wenn Sie wie die meisten anderen Menschen sind, entscheiden Sie sich bei dieser Frage für die Lotterie. Die Erklärung für diese risikoreiche Wahl ist das Spiegelbild der Erklärung für die Risikoaversion bei Problem 1: Der (negative) Wert eines Verlusts von 900 Dollar ist viel größer als 90 Prozent des (negativen) Werts eines Verlusts von 1000 Dollar. Der sichere Verlust ruft eine sehr starke Risikovermeidung hervor, und dies veranlasst einen dazu, das Risiko einzugehen. Später werden wir sehen, dass die Beurteilungen der Wahrscheinlichkeiten (90 gegen 100 Prozent) ebenfalls sowohl zu der Risikoscheu bei Problem 1 als auch zu der Präferenz für das Glücksspiel bei Problem 2 beitragen.

Wir haben nicht als Erste bemerkt, dass Menschen risikofreudig werden, wenn all ihre Optionen negativ sind, aber die theorieinduzierte Blindheit hatte sich durchgesetzt. Weil die vorherrschende Theorie keine plausible Möglichkeit bot, verschiedenen Einstellungen zu Gewinn- und Verlustrisiken Rechnung zu tragen, musste die Tatsache, dass sich die Einstellungen unterschieden, ignoriert werden. Dagegen veranlasste uns die

Entscheidung, Ergebnisse als Gewinne und Verluste zu betrachten, dazu, uns genau auf diese Diskrepanz zu konzentrieren.

Die Beobachtung gegensätzlicher Einstellungen zu Risiken mit günstigen und ungünstigen Aussichten brachte uns schon bald einen bedeutenden Schritt weiter: Wir fanden einen Weg, um den zentralen Irrtum in Bernoullis Entscheidungsmodell nachzuweisen. Betrachten Sie folgende Fallbeispiele:

> Problem 3: Zusätzlich zu Ihrem bisherigen Vermögen erhalten Sie
> 1000 Dollar. Jetzt sollen Sie sich für eine dieser Optionen entscheiden:
> Eine 50-prozentige Chance, 1000 Dollar zu gewinnen,
> oder 500 Dollar sicher erhalten.

> Problem 4: Zusätzlich zu Ihrem bisherigen Vermögen erhalten Sie
> 2000 Dollar. Jetzt sollen Sie sich für eine dieser Optionen entscheiden:
> Eine 50-prozentige Wahrscheinlichkeit, 1000 Dollar zu verlieren,
> oder 500 Dollar sicher verlieren.

Sie können leicht überprüfen, dass im Hinblick auf die endgültigen Vermögenszustände – alles, was für Bernoullis Theorie von Belang ist – Problem 3 und 4 identisch sind. In beiden Fällen hat man die Wahl zwischen zwei gleichen Optionen: Entweder man ist mit Sicherheit um 1500 Dollar reicher, als man es gegenwärtig ist, oder man akzeptiert ein Glücksspiel, bei dem man mit 50-prozentiger Wahrscheinlichkeit entweder um 1000 Dollar oder um 2000 Dollar reicher sein wird. Daher sollten die beiden Probleme gemäß Bernoullis Theorie ähnliche Präferenzen erzeugen. Überprüfen Sie Ihre Intuitionen, und Sie werden vermutlich ahnen, wie sich andere Menschen entschieden haben.

– Bei der ersten Wahl zog eine große Mehrheit der Befragten die sichere Option vor.
– Bei der zweiten Wahl zog eine große Mehrheit die Lotterie vor.

Das Finden unterschiedlicher Präferenzen bei den Problemen 3 und 4 war ein maßgebliches Gegenbeispiel gegen die Schlüsselidee von Bernoullis Theorie. Wenn es allein auf den Nutzen des Vermögens ankäme, dann sollten die offensichtlich äquivalenten Formulierungen desselben Problems zu den gleichen Entscheidungen führen. Der Vergleich der Probleme verdeutlicht die alles entscheidende Rolle des Referenzpunktes, von dem aus die Optionen bewertet werden. Bei Problem 3 liegt der Referenzpunkt um 1000 Dollar über dem gegenwärtigen Vermögen, bei Problem 4 um 2000 Dollar. Eine Bereicherung um 1500 Dollar entspricht daher bei Problem 3 einem Gewinn von 500 Dollar und bei Problem 4 einem Verlust. Weitere Beispiele der gleichen Art lassen sich offenkundig leicht erzeugen. Die Geschichte von Anthony und Betty hatte eine ähnliche Struktur.

Wie viel Aufmerksamkeit widmeten Sie dem Geschenk von 1000 oder 2000 Dollar, das Ihnen »gemacht« wurde, ehe Sie Ihre Wahl trafen? Wenn Sie wie die meisten Menschen sind, haben Sie kaum Notiz davon genommen. Tatsächlich gab es keinen Grund für Sie, es weiter zu beachten, weil das Geschenk in den Referenzpunkt einbezogen ist, und Referenzpunkte werden im Allgemeinen ignoriert. Sie wissen etwas über Ihre Präferenzen, was Nutzentheoretiker nicht wissen – dass sich Ihre Einstellung zu Risiken nicht verändern würde, wenn Ihr Nettovermögen um ein paar Tausend Dollar höher oder niedriger läge (es sei denn, Sie wären sehr arm). Und Sie wissen auch, dass Ihre Einstellungen zu Gewinnen und Verlusten nicht von Ihrer Beurteilung Ihrer Vermögenslage abhängig sind. Der Grund dafür, dass Ihnen der Gedanke, 100 Dollar zu gewinnen, gefällt, während Ihnen der Gedanke, 100 Dollar zu verlieren, missfällt, besteht nicht darin, dass diese Summen Ihr Vermögen verändern. Sie gewinnen einfach gern und verlieren ungern – und höchstwahrscheinlich ist Ihre Abneigung dagegen, etwas zu verlieren, stärker als Ihre Neigung, etwas zu gewinnen.

Die vier Probleme verdeutlichen die Schwäche von Bernoullis Modell. Seine Theorie ist zu einfach, und ihr fehlt ein bewegliches Teil. Die fehlende Variable ist der Referenzpunkt, der frühere Zustand, auf den sich die Beurteilung von Gewinnen und Verlusten bezieht. In Bernoullis Theorie muss man nur den Vermögenszustand kennen, um dessen Nutzen

bestimmen zu können, in der Neuen Erwartungstheorie hingegen muss man auch den Referenzzustand kennen. Daher ist die Neue Erwartungstheorie komplexer als die Nutzentheorie. In der Wissenschaft gilt Komplexität als ein Kostenfaktor, der durch eine hinreichend vielfältige Menge neuer und (vorzugsweise) interessanter Vorhersagen von Tatsachen, die die gegenwärtige Theorie nicht erklären kann, gerechtfertigt werden muss. Dies war die Herausforderung, vor der wir standen.

Obgleich Amos und ich nicht mit dem kognitionspsychologischen Zwei-Systeme-Modell arbeiteten, steht heute fest, dass im Zentrum der Neuen Erwartungstheorie drei kognitive Merkmale stehen. Sie spielen eine wesentliche Rolle bei der Bewertung finanzieller Ergebnisse und sind vielen automatischen Wahrnehmungs-, Urteils- und emotionalen Prozessen gemeinsam. Sie sollten als operative Merkmale von System 1 betrachtet werden.

– Die Beurteilung erfolgt in Bezug auf einen neutralen Referenzpunkt, der manchmal auch als »Anpassungsniveau« bezeichnet wird. Dieses Prinzip lässt sich mithilfe einer einfachen Demonstration sehr gut veranschaulichen. Stellen Sie drei Schüsseln vor sich auf einen Tisch. Füllen Sie die linke Schüssel mit eiskaltem und die rechte Schüssel mit warmem Wasser. Das Wasser in der mittleren Schüssel sollte Raumtemperatur haben. Tauchen Sie eine Hand für etwa eine Minute in das kalte Wasser und die andere ins warme Wasser, tauchen Sie dann beide in die mittlere Schüssel. Sie werden dieselbe Temperatur in der einen Hand als warm und in der anderen Hand als kalt empfinden. Für finanzielle Ergebnisse ist der übliche Referenzpunkt der Status quo. Aber er kann auch das Ergebnis sein, das Sie erwarten, oder vielleicht das Ergebnis, das Ihnen Ihres Erachtens zusteht, zum Beispiel die Gehaltserhöhung oder die Prämie, die Ihre Kollegen erhalten. Ergebnisse, die besser sind als die Referenzpunkte, sind Gewinne. Unterhalb des Referenzpunktes sind sie Verluste.
– Ein Prinzip abnehmender Empfindlichkeit gilt sowohl für Wahrnehmungsdimensionen als auch für die Beurteilung von Vermögensänderungen. Das Einschalten eines schwachen Lichts hat in einem dunklen Raum eine starke Wirkung. Der gleiche Helligkeitszuwachs in einem hell erleuchteten Raum mag nicht wahrnehmbar sein. In ähnlicher Weise ist die subjektive Differenz

zwischen 900 und 1000 Dollar viel kleiner als die Differenz zwischen 100 und 200 Dollar.

– Das dritte Prinzip ist die Verlustaversion. Wenn sie direkt miteinander verglichen oder gegeneinander gewichtet werden, schlagen Verluste stärker zu Buche als Gewinne. Diese Asymmetrie zwischen der Macht positiver und negativer Erwartungen oder Erfahrungen ist evolutionsgeschichtlich bedingt. Lebewesen, die Bedrohungen vordringlicher behandeln als Chancen, haben höhere Überlebens- und Fortpflanzungschancen.

Die drei Prinzipien, die den Wert von Ergebnissen bestimmen, werden in Abbildung 10 veranschaulicht.[2] Wenn die Neue Erwartungstheorie eine Flagge hätte, wäre sie mit diesem Bild bemalt. Der Graph zeigt den psychologischen Wert von Gewinnen und Verlusten, die in der Neuen Erwartungstheorie die »Träger« des Wertes sind (anders als in Bernoullis Modell, wo Vermögenszustände die Träger des Wertes sind). Der Graph hat zwei verschiedene Teile, rechts und links eines neutralen Referenzpunktes. Ein markantes Merkmal ist der s-förmige Verlauf, der die abnehmende Empfindlichkeit für Gewinne und Verluste repräsentiert. Schließlich sind die beiden Kurven des S nicht symmetrisch. Die Steigung der Funktion ändert sich plötzlich im Referenzpunkt: Die Reaktion auf Verluste ist stärker als die Reaktion auf entsprechende Gewinne. Dies ist die Verlustaversion.

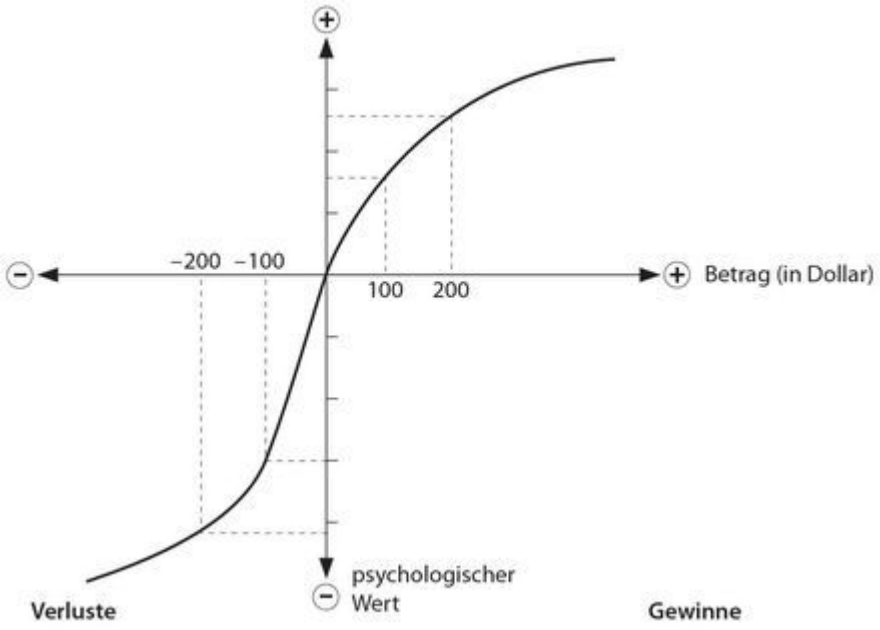


**Abbildung 10**

## Verlustaversion

Viele der Optionen, denen wir uns im Leben gegenübersehen, sind »gemischt«: Es gibt ein Verlustrisiko und eine Gewinnchance, und wir müssen entscheiden, ob wir das Risiko eingehen oder nicht. Investoren, die ein Start-up-Unternehmen beurteilen, Anwälte, die sich fragen, ob sie prozessieren sollen, Generäle, die in Kriegszeiten eine Offensive in Betracht ziehen, und Politiker, die entscheiden müssen, ob sie sich um ein Amt bewerben sollen, sind alle mit den Möglichkeiten von Sieg und Niederlage konfrontiert. Für ein elementares Beispiel gemischter Aussichten betrachten Sie Ihre Reaktion auf die nächste Frage.

> Problem 5: Man bietet Ihnen eine Lotterie auf einen Münzwurf an.
> Wenn die Münze Zahl zeigt, verlieren Sie 100 Dollar.
> Wenn die Münze Kopf zeigt, gewinnen Sie 150 Dollar.
> Ist diese Wette attraktiv? Würden Sie darauf eingehen?

Um diese Entscheidung zu treffen, müssen Sie den psychologischen Nutzen eines Gewinns von 150 Dollar gegen die psychologischen Kosten eines Verlusts von 100 Dollar abwägen. Was halten Sie davon? Obgleich der erwartete Nutzen der Lotterie offensichtlich positiv ist, weil Sie mehr gewinnen als verlieren können, lehnen Sie diese wahrscheinlich ab – die meisten Menschen tun das. Die Ablehnung dieses Glücksspiels geht auf das Konto von System 2, aber die entscheidenden Inputs sind emotionale Reaktionen, die von System 1 generiert werden. Für die meisten Menschen ist die Furcht, 100 Dollar zu verlieren, stärker als die Hoffnung, 150 Dollar zu gewinnen. Aus vielen derartigen Beobachtungen folgerten wir, dass Verluste höher bewertet werden als Gewinne und dass Menschen verlustscheu sind.

Sie können das Ausmaß Ihrer Verlustaversion dadurch messen, dass Sie sich selbst eine Frage stellen: Was ist der niedrigste Gewinn, den ich brauche, um die 50-prozentige Wahrscheinlichkeit eines Verlusts von 100

Dollar auszugleichen? Für viele Menschen liegt die Antwort bei etwa 200 Dollar, also der doppelten Höhe des Verlusts. Die Verlustaversionsrate liegt, wie in mehreren Experimenten nachgewiesen wurde, zwischen 1,5 und 2,5.[3] Dies ist selbstverständlich ein Mittelwert; einige Menschen haben eine viel stärkere Verlustaversion als andere. Professionelle Risiko-Akteure an den Finanzmärkten sind toleranter gegenüber Verlusten, vermutlich weil sie nicht auf jede Schwankung emotional reagieren. Als die Teilnehmer eines Experiments »wie ein Wertpapierhändler denken« sollten, nahm ihre Verlustaversion ab, und ihre emotionale Reaktion auf Verluste (gemessen durch einen physiologischen Index für emotionale Erregung) ging deutlich zurück. [4]

Um Ihre Verlustaversionsrate für verschiedene Einsätze zu überprüfen, sollten Sie die folgenden Fragen betrachten. Klammern Sie alle sozialen Erwägungen aus, versuchen Sie nicht, entweder wagemutig oder vorsichtig zu erscheinen, und konzentrieren Sie sich allein auf die subjektive Auswirkung des möglichen Verlusts und des kompensierenden Gewinns.

– Betrachten Sie eine 50:50-Lotterie, bei der Sie 10 Dollar verlieren können. Was ist der kleinste Gewinn, der die Wette attraktiv macht? Wenn Sie 10 Dollar sagen, sind Sie gleichgültig gegenüber dem Risiko. Wenn Sie eine Zahl von unter 10 Dollar angeben, sind Sie risikofreudig. Wenn Ihre Antwort über 10 Dollar liegt, sind Sie verlustscheu.
– Wie steht es mit einem möglichen Verlust von 500 Dollar bei einem Münzwurf? Welchen potenziellen Gewinn verlangen Sie, um diesen auszugleichen?
– Wie verhält es sich bei einem möglichen Verlust von 2000 Dollar?

Sie stellten bei dieser Übung vermutlich fest, dass Ihr Verlustaversionskoeffizient mit steigenden Einsätzen tendenziell ansteigt, wenn auch nicht dramatisch. Natürlich »geht gar nichts«, wenn der mögliche Verlust potenziell ruinös ist oder wenn Ihr Lebensstil bedroht ist. In solchen Fällen ist der Verlustaversionskoeffizient sehr groß und vielleicht sogar unendlich – es gibt Risiken, die Sie einfach nicht eingehen werden, ganz egal wie viele Millionen Sie gewinnen könnten, wenn Sie Glück hätten.

Ein erneuter Blick auf Abbildung 10 mag helfen, einem weitverbreiteten Missverständnis vorzubeugen. In diesem Kapitel habe ich zwei Behauptungen aufgestellt, die einigen Lesern widersprüchlich erscheinen mögen:

– In gemischten Lotterien, bei denen sowohl ein Gewinn als auch ein Verlust möglich ist, führt die Verlustaversion zu extrem risikoscheuen Entscheidungen.
– Bei schlechten Wahlmöglichkeiten, bei denen ein sicherer Verlust mit einem höheren Verlust, der lediglich wahrscheinlich ist, verglichen wird, führt rückläufige Empfindlichkeit zu Risikofreude.

Das ist kein Widerspruch. Im gemischten Fall wird der potenzielle Verlust doppelt so hoch bewertet wie der potenzielle Gewinn, wie leicht ersichtlich ist, wenn man die Steigungen der Wertfunktion für Verluste und Gewinne vergleicht. Bei der »Wahl zwischen zwei Übeln« verursacht die Krümmung der Wertkurve (abnehmende Sensibilität) Risikofreude. Das Unbehagen über einen Verlust von 900 Dollar wiegt schwerer als 90 Prozent des Unbehagens über einen Verlust von 1000 Dollar. Diese beiden Erkenntnisse sind die Quintessenz der Neuen Erwartungstheorie.

Abbildung 10 zeigt eine plötzliche Veränderung in der Steigung der Wertfunktion dort, wo Gewinne zu Verlusten werden, weil es auch dann, wenn der auf dem Spiel stehende Betrag im Verhältnis zu Ihrem Vermögen verschwindend gering ist, eine erhebliche Verlustabneigung gibt. Ist es plausibel, dass Einstellungen zu Vermögenszuständen die extreme Abneigung gegenüber geringen Risiken erklären könnten? Es ist ein bemerkenswertes Beispiel für theorieinduzierte Blindheit, dass dieser offenkundige Fehler in Bernoullis Theorie 250 Jahre lang von der Wissenschaft unbemerkt blieb. Im Jahr 2000 erbrachte der Verhaltensökonom Matthew Rabin endlich den mathematischen Beweis, dass Versuche, die Verlustaversion mit dem Nutzen des Vermögens zu erklären, absurd und zum Scheitern verurteilt sind – und sein Beweis fand große Beachtung. Rabins Theorem zeigt, dass jeder, der eine günstige Lotterie mit geringen Einsätzen ablehnt, aus mathematischen Gründen auch eine absurd hohe Risikoaversion für eine Wette mit höheren Einsätzen hätte.[5] Er weist zum Beispiel darauf hin, dass die meisten *Humans* die folgende Lotterie ablehnen:

> 50-prozentige Chance, 100 Dollar zu verlieren, und 50-prozentige Chance, 200 Dollar zu gewinnen.

Dann zeigt er, dass eine Person, die dieses Glücksspiel ablehnt, laut der Nutzentheorie auch die folgende Lotterie ablehnen wird:

> 50-prozentige Chance, 200 Dollar zu verlieren, und 50-prozentige Chance, 20 000 Dollar zu gewinnen.

Selbstverständlich wird niemand, der bei klarem Verstand ist, diese Lotterie ablehnen![6] In einem überschwänglichen Artikel, den sie über den Beweis schrieben, bemerkten Matthew Rabin und Richard Thaler, die Lotterie mit höherem Einsatz »hat einen erwarteten Ertrag von 9900 Dollar – und eine Wahrscheinlichkeit von genau null, mehr als 200 Dollar zu verlieren. Selbst ein lausiger Anwalt könnte einen dafür, dass man dieses Glücksspiel ausschlägt, für unzurechnungsfähig erklären lassen.«[7] Vielleicht haben sie im Schwung ihrer Begeisterung ihren Aufsatz mit einem Verweis auf einen berühmten Sketch von Monty Python beendet, in dem ein frustrierter Kunde versucht, einen toten Papagei in einer Tierhandlung zurückzugeben. Der Kunde beschreibt wortreich den Zustand des Papageis, und seine Ausführungen gipfeln in der Aussage: »Dies ist ein Expapagei.« Rabin und Thaler schlossen daran die Bemerkung an: »Es ist Zeit für die Wirtschaftswissenschaftler, anzuerkennen, dass der Erwartungsnutzen eine Exhypothese ist.« Viele Ökonomen empfanden diese respektlose Aussage beinahe als eine Art Gotteslästerung. Doch die theorieinduzierte Blindheit, die sich darin zeigt, den Nutzen des Vermögens als eine Erklärung für Einstellungen gegenüber kleinen Verlusten zu akzeptieren, ist eine legitime Zielscheibe für humorvolle Kommentare.

## Blinde Flecken der Neuen Erwartungstheorie

Bislang habe ich die Vorzüge der Neuen Erwartungstheorie gelobt und das rationale Modell und die Erwartungsnutzentheorie kritisiert. Es ist Zeit für eine gewisse Ausgewogenheit.

Die meisten Studenten der Volkswirtschaftslehre haben von der Neuen Erwartungstheorie und der Verlustaversion gehört, aber diese Begriffe sucht man in der Regel vergeblich im Register einführender Lehrbücher dieser Disziplin. Diese Auslassung schmerzt mich manchmal, aber tatsächlich ist sie durchaus verständlich, wenn man an die zentrale Rolle der Rationalität in der ökonomischen Standardtheorie denkt. Die gängigen Konzepte und Ergebnisse, die Studenten im Grundstudium beigebracht werden, lassen sich leichter erklären, wenn man annimmt, dass *Econs* keine törichten Fehler machen. Diese Annahme ist wirklich notwendig, und sie würde untergraben, wenn man die *Humans* der Neuen Erwartungstheorie einführte, deren Beurteilungen von Ergebnissen unangemessen kurzsichtig sind.

Es gibt gute Gründe dafür, die Neue Erwartungstheorie in einführenden Lehrbüchern nicht zu behandeln. Die Grundbegriffe der Volkswirtschaftslehre sind anspruchsvolle und leistungsfähige intellektuelle Werkzeuge, die selbst bei vereinfachten und unrealistischen Annahmen über die Natur ökonomischer Akteure, die auf Märkten miteinander interagieren, nicht leicht zu verstehen sind. Es wäre verwirrend und vielleicht sogar entmutigend, diese Annahmen infrage zu stellen. Es ist vernünftig, den Studenten zunächst einmal dabei zu helfen, sich die grundlegenden Werkzeuge dieses Fachgebiets anzueignen. Außerdem ist das Versagen der Rationalität, das in die Neue Erwartungstheorie eingebaut ist, oftmals irrelevant für die Vorhersagen der ökonomischen Theorie, die in manchen Situationen sehr präzise gelingen und in vielen anderen gute Näherungen liefern. In einigen Kontexten ist der Unterschied jedoch von Bedeutung: Das Verhalten der von der Neuen Erwartungstheorie beschriebenen *Humans* wird von der sofortigen emotionalen Wirkung von Gewinnen und Verlusten geleitet, nicht von langfristigen Vermögensaussichten und der globalen Nutzenerwartung.

In meiner Diskussion von Mängeln in Bernoullis Modell, das über 200 Jahre lang nicht hinterfragt wurde, betonte ich das Phänomen der theorieinduzierten Blindheit. Aber natürlich ist die theorieinduzierte Blindheit nicht auf die Erwartungsnutzentheorie beschränkt. Die Neue Erwartungstheorie hat ebenfalls ihre Mängel, und die theorieinduzierte Blindheit für diese Mängel hat dazu beigetragen, dass sie als hauptsächliche Alternative für die Nutzentheorie anerkannt wurde.

Betrachten wir die Annahme der Neuen Erwartungstheorie, dass der Referenzpunkt, im Allgemeinen der Status quo, einen Wert von null hat. Diese Annahme scheint vernünftig zu sein, aber sie führt zu einigen absurden Konsequenzen. Betrachten Sie genau die folgenden Gewinnerwartungen. Wie sind sie zu beurteilen?

a. Eine Chance von eins zu einer Million, 1 Million Dollar zu gewinnen.
b. Eine 90-prozentige Chance, 12 Dollar zu gewinnen, und eine 10-prozentige Chance, nichts zu gewinnen.
c. Eine 90-prozentige Chance, 1 Million Dollar zu gewinnen, und eine 10-prozentige Chance, nichts zu gewinnen.

Nichts zu gewinnen ist bei allen drei Glücksspielen ein mögliches Resultat, und die Neue Erwartungstheorie weist diesem Ergebnis in den drei Fällen den gleichen Wert zu. Nichts zu gewinnen ist der Referenzpunkt, und sein Wert ist null. Entsprechen diese Aussagen unserer Erfahrung? Natürlich nicht. Nichts zu gewinnen ist in den ersten beiden Fällen ein Nichtereignis, und es ist daher durchaus sinnvoll, ihm einen Wert von null zuzuschreiben. Im dritten Szenario ist es äußerst enttäuschend, nichts zu gewinnen. Wie eine informell versprochene Gehaltserhöhung setzt die hohe Wahrscheinlichkeit, die große Summe zu gewinnen, einen vorläufigen neuen Referenzpunkt fest. Vor dem Hintergrund der subjektiven Erwartungen erleben wir es als einen großen Verlust, nichts zu gewinnen. Die Neue Erwartungstheorie kann dieser Tatsache nicht Rechnung tragen, weil sie es nicht zulässt, dass sich der Wert eines Ergebnisses (in diesem Fall: nichts zu gewinnen) verändert, wenn es hochgradig unwahrscheinlich oder die Alternative sehr hochwertig ist. Kurz gesagt, die Neue Erwartungstheorie kann Enttäuschung nicht abbilden. Enttäuschung und die Erwartung von Enttäuschung sind jedoch real, und wenn man sie nicht berücksichtigt, ist dies genauso ein offensichtlicher Mangel wie die

Gegenbeispiele, die ich angeführt habe, um Bernoullis Theorie zu kritisieren.

Auch Bereuen wird von der Neuen Erwartungstheorie und der Nutzentheorie nicht berücksichtigt. Die beiden Theorien teilen die Annahme, dass verfügbare Optionen bei einer Wahl getrennt und unabhängig voneinander beurteilt werden und dass die Option mit dem höchsten Wert ausgewählt wird. Diese Annahme ist mit Sicherheit falsch, wie das folgende Beispiel zeigt.

> Problem 6: Wählen Sie zwischen einer 90-prozentigen Chance, 1 Million Dollar zu gewinnen, und zu 100 Prozent sicheren 50 Dollar.

> Problem 7: Wählen Sie zwischen einer 90-prozentigen Chance, 1 Million Dollar zu gewinnen, und zu 100 Prozent sicheren 150 000 Dollar.

Vergleichen Sie den vorweggenommenen seelischen Schmerz, sich in beiden Fällen für die Lotterie zu entscheiden, aber *nicht* zu gewinnen. In beiden Fällen ist es eine Enttäuschung, nichts zu gewinnen, aber in Problem 7 wird der potenzielle Schmerz durch das Wissen verschlimmert, dass man die »habgierige« Entscheidung bereuen wird, wenn man sich für das Glücksspiel entscheidet und verliert, da man ein sicheres Geschenk von 150 000 Dollar ausschlug. Man bereut ein Ergebnis dann, wenn man eine hochwertige Option, die man hätte wählen können, verworfen hat.

Mehrere Ökonomen und Psychologen haben Modelle der Entscheidungsfindung vorgeschlagen, die auf den Emotionen von Reue und Enttäuschung basieren.[8] Es trifft wohl zu, dass diese Modelle weniger einflussreich waren als die Neue Erwartungstheorie, und der Grund dafür ist aufschlussreich. Die Emotionen von Reue und Enttäuschung sind real, und Entscheider antizipieren bei der Entscheidungsfindung sicherlich diese Gefühle. Das Problem besteht darin, dass Theorien auf der Basis von Bedauern nur wenige treffende Vorhersagen machen, die sie gegenüber der

Neuen Erwartungstheorie auszeichnen würden, welche den Vorteil hat, einfacher zu sein. Die Komplexität der Neuen Erwartungstheorie war im Wettstreit mit der Erwartungsnutzentheorie akzeptabler, weil sie Beobachtungen vorhersagte, die die Erwartungsnutzentheorie nicht erklären konnte.

Differenziertere und realistischere Annahmen genügen nicht, um eine Theorie erfolgreich zu machen. Wissenschaftler benutzen Theorien als einen Werkzeugkasten, und sie belasten sich nur dann mit einem schwereren Kasten, wenn die neuen Werkzeuge sehr nützlich sind. Die Neue Erwartungstheorie wurde von vielen Wissenschaftlern nicht deshalb anerkannt, weil sie »wahr« ist, sondern weil die Konzepte, die sie zusätzlich in die Nutzentheorie einführte, insbesondere den Referenzpunkt und die Verlustaversion, die Mühe wert waren; sie erbrachten neue Vorhersagen, die sich als zutreffend erwiesen. Wir hatten Glück.

## Zum Thema »Neue Erwartungstheorie«

»Er leidet an einer extremen Verlustaversion, die ihn dazu veranlasst, alle günstigen Chancen zu verwerfen.«

»In Anbetracht ihres riesigen Vermögens ist ihre emotionale Reaktion auf triviale Gewinne und Verluste unverständlich.«

»Er gewichtet Verluste etwa doppelt so hoch wie Gewinne, was normal ist.«

# *27. Der Endowment-Effekt*

Vermutlich haben Sie das Diagramm in Abbildung 11 – oder ein ähnliches – schon einmal irgendwo gesehen, auch wenn Sie keinen Unterricht in Wirtschaftswissenschaften hatten. Der Graph zeigt das sogenannte »Indifferenzkurvensystem« eines Individuums für zwei Güter.

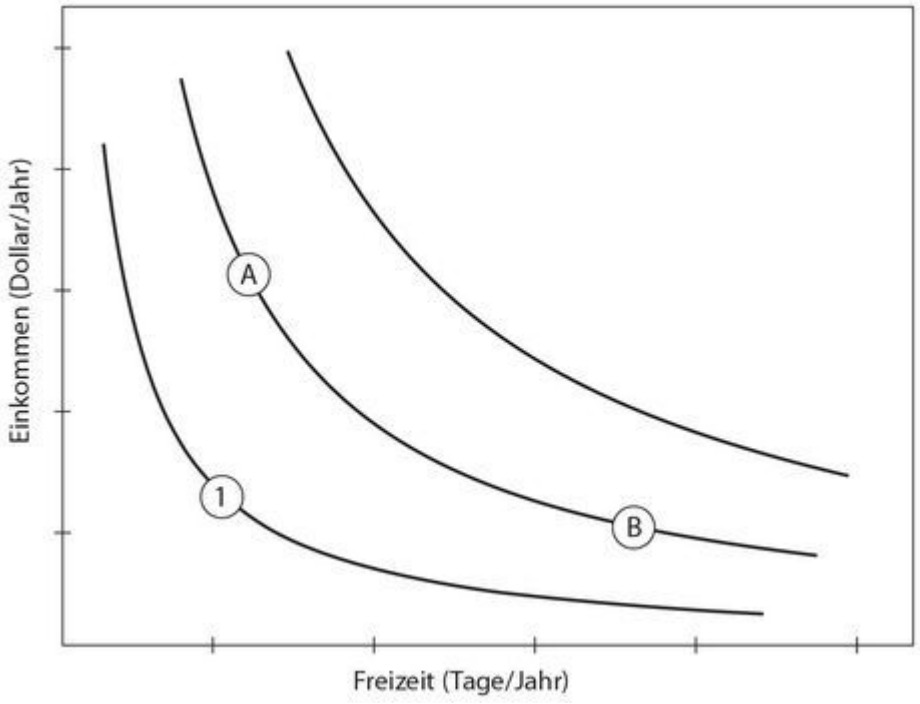


**Abbildung 11**

In Einführungskursen in die Volkswirtschaftslehre lernen Studenten, dass jeder Punkt auf den Kurven eine bestimmte Kombination von Einkommen und Anzahl der Urlaubstage spezifiziert. Jede »Indifferenzkurve« verknüpft die Kombinationen der beiden Güter, die gleich begehrenswert sind – den gleichen Nutzen haben. Die Kurven würden zu parallelen Geraden, wenn Menschen bereit wären, Urlaubstage zum selben Preis für zusätzliches Einkommen zu »verkaufen«, unabhängig davon, wie viel Einkommen und wie viel Urlaubszeit sie haben. Die konvexe Form veranschaulicht den rückläufigen Grenznutzen: Je mehr Freizeit man hat, umso weniger wert ist einem ein zusätzlicher Urlaubstag, und jeder zusätzliche Tag ist weniger

wert als der Tag zuvor. In ähnlicher Weise ist es so, dass mit steigendem Einkommen der Wert jedes zusätzlichen Dollars abnimmt, während der Betrag ansteigt, auf den man für einen zusätzlichen Urlaubstag zu verzichten bereit ist.

Alle Positionen auf einer Indifferenzkurve sind gleichermaßen attraktiv. Genau das bedeutet »Indifferenz«: Es ist einem egal, wo man sich auf einer Indifferenzkurve befindet. Wenn A und B für Sie auf derselben Indifferenzkurve liegen, sind Sie ihnen gegenüber indifferent, und Sie brauchen keinen Anreiz, um sich von einem Punkt zum anderen oder zurück zu bewegen. Eine Version dieses Diagramms findet sich in jedem wirtschaftswissenschaftlichen Lehrbuch der letzten hundert Jahre, und viele Millionen Studenten haben diese Kurven betrachtet. Aber nur wenigen ist aufgefallen, was fehlt. Die Erklärungskraft und Eleganz eines theoretischen Modells haben Studenten und Wissenschaftler ein weiteres Mal blind gemacht für einen gravierenden Mangel.

In der Abbildung fehlt nämlich die Angabe des gegenwärtigen Einkommens und der gegenwärtigen Urlaubstage eines Individuums.[1] Wenn Sie Angestellter sind, sind in Ihrem Arbeitsvertrag die Höhe Ihres Gehalts und die Anzahl Ihrer Urlaubstage genau festgelegt – und diese entsprechen einem Punkt auf einer Kurve. Dies ist Ihr Referenzpunkt, Ihr Status quo, aber das Diagramm zeigt diesen nicht. Durch das Weglassen dieses Punktes wiegen die Theoretiker, die dieses Diagramm zeichnen, Sie in dem Glauben, der Referenzpunkt sei irrelevant, aber Sie wissen jetzt, dass er natürlich von Belang ist. Da haben wir ein weiteres Mal Bernoullis Irrtum. Die Darstellung der Indifferenzkurven geht stillschweigend davon aus, dass Ihr Nutzen zu jedem beliebigen Zeitpunkt ausschließlich von Ihrer gegenwärtigen Situation bestimmt wird, dass die Vergangenheit irrelevant ist und dass Ihre Beurteilung eines potenziellen Arbeitsplatzes nicht von den Bedingungen Ihres gegenwärtigen Arbeitsvertrags abhängig ist. Diese Annahmen sind in diesem Fall und in vielen weiteren Fällen vollkommen unrealistisch.

Das Weglassen des Referenzpunktes im Indifferenzkurvensystem ist ein überraschender Fall von theorieinduzierter Blindheit, weil wir sehr oft Fällen begegnen, in denen der Referenzpunkt offenkundig von Bedeutung ist. Bei Lohn- und Tarifverhandlungen wissen beide Seiten ganz genau, dass der bestehende Vertrag der Referenzpunkt ist und dass sich die Verhandlungen auf wechselseitige Forderungen nach Zugeständnissen in

Bezug auf diesen Referenzpunkt konzentrieren werden. Die Rolle der Verlustaversion bei Verhandlungen ist ebenfalls allgemein bekannt: Es schmerzt, Zugeständnisse zu machen. Sie haben reichlich persönliche Erfahrungen mit der Rolle des Referenzpunktes. Wenn Sie Ihren Arbeitsplatz oder Ihren Wohnort wechselten oder auch nur einen solchen Wechsel in Betracht zogen, erinnern Sie sich bestimmt, dass Sie die Merkmale der neuen Stelle oder Umgebung als Plus- oder Minuspunkte im Vergleich zu Ihrem Status quo bewerteten. Vielleicht haben Sie auch bemerkt, dass bei dieser Beurteilung Nachteile stärker zu Buche schlugen als Vorteile – die Verlustaversion war am Werk. Es ist schwer, Veränderungen zum Schlechteren zu akzeptieren. So beträgt zum Beispiel der Mindestlohn, den Arbeitslose bei einem neuen Beschäftigungsverhältnis akzeptieren würden, im Durchschnitt 90 Prozent ihres letzten Lohns, und dieser Betrag sinkt über einen einjährigen Zeitraum um weniger als 10 Prozent.[2]

Um eine Vorstellung davon zu bekommen, wie stark sich der Referenzpunkt auf Entscheidungen auswirkt, wollen wir uns das Beispiel von Albert und Ben ansehen, »hedonistischen Zwillingen«, die gleiche Präferenzen und gleiche Ausgangsjobs mit niedrigem Einkommen und wenig Urlaub haben. Ihre gegenwärtige Lage entspricht dem mit 1 gekennzeichneten Punkt in Abbildung 11. Die Firma bietet ihnen zwei bessere Stellen an und lässt sie entscheiden, wer eine Lohnerhöhung von 10 000 Dollar (Position A) und wer jeden Monat einen zusätzlichen Tag bezahlten Urlaub (Position B) bekommt. Da beide indifferent sind, werfen sie eine Münze. Albert entscheidet sich daraufhin für die Lohnerhöhung, Ben für den zusätzlichen Urlaub. Es vergeht einige Zeit, in der sich die Zwillinge an ihre neuen Arbeitsplätze gewöhnen. Dann macht ihnen die Firma das Angebot, die Stellen zu tauschen, wenn sie wollen.

Die in der Abbildung dargestellte Standardtheorie geht davon aus, dass Präferenzen im Zeitablauf stabil sind. Die Positionen A und B sind für beide Zwillinge gleich attraktiv, und sie brauchen nur einen geringen oder gar keinen Anreiz, um zu tauschen. In scharfem Gegensatz dazu behauptet die Neue Erwartungstheorie, dass beide Zwillinge auf jeden Fall vorziehen werden, auf ihren jeweiligen Positionen zu bleiben. Diese Präferenz für den Status quo ist eine Folge der Verlustaversion.

Konzentrieren wir uns auf Albert. Er befand sich anfänglich in Position 1 auf dem Graphen, und von diesem Referenzpunkt aus fand er diese beiden

Alternativen gleich attraktiv:

Gehe nach A: Eine Erhöhung um 10 000 Dollar.

Oder:

Gehe nach B: Zwölf zusätzliche Urlaubstage.

Der Wechsel zur Position A verändert Alberts Referenzpunkt, und wenn er erwägt, nach B zu wechseln, hat seine Wahl eine neue Struktur:

Bleib auf A: Kein Gewinn und kein Verlust.

Oder:

Wechsle zu B: Zwölf zusätzliche Urlaubstage und eine Gehaltskürzung von 10 000 Dollar.

Sie haben gerade die subjektive Erfahrung der Verlustaversion gemacht. Sie konnten es fühlen: Eine Gehaltskürzung von 10 000 Dollar ist eine sehr schlechte Nachricht. Selbst wenn ein Gewinn von zwölf Urlaubstagen genauso eindrucksvoll war wie ein Gewinn von 10 000 Dollar, ist die gleiche Verlängerung der Urlaubszeit nicht ausreichend, um einen Verlust von 10 000 Dollar zu kompensieren. Albert wird auf A bleiben, weil der Nachteil eines Wechsels den Vorteil überwiegt. Die gleiche Überlegung gilt für Ben, der ebenfalls seinen gegenwärtigen Arbeitsplatz behalten will, weil der Verlust der mittlerweile hochgeschätzten zusätzlichen Urlaubstage schwerer wiegt als der Nutzen des zusätzlichen Einkommens.

Dieses Beispiel verdeutlicht zwei Aspekte von Wahlhandlungen, die das Standardmodell der Indifferenzkurven nicht vorhersagt. Erstens, Präferenzen sind nicht konstant; sie schwanken mit dem Referenzpunkt. Zweitens, die Nachteile eines Wechsels fallen stärker ins Gewicht als seine Vorteile, was eine Verzerrung erzeugt, die den Status quo begünstigt.

Natürlich folgt aus der Verlustaversion nicht, dass man es niemals vorzieht, seine Situation zu verändern; die Vorteile einer Chance mögen sogar schwerer wiegen als übergewichtete Verluste. Aus der Verlustaversion folgt lediglich, dass Wahlen eine starke Verzerrung zugunsten der Referenzsituation aufweisen (und eine allgemeine Tendenz zur Begünstigung kleiner statt großer Veränderungen zeigen).

Herkömmlichen Indifferenzkurven und Bernoullis Darstellung von Ergebnissen als Vermögenszuständen ist eine Fehlannahme gemeinsam: dass der individuelle Nutzen eines Zustands nur von diesem Zustand abhängig sei und nicht von der individuellen Geschichte beeinflusst würde. Es war eine der Leistungen der Verhaltensökonomik, dass sie diesen Fehler korrigiert hat.

## Der Endowment-Effekt

Die Frage, wann eine bestimmte Theorie oder eine Bewegung ihren Anfang nahm, ist oftmals schwer zu beantworten, aber der Ursprung dessen, was heute »Verhaltensökonomik« genannt wird, lässt sich genau angeben. Anfang der 1970er-Jahre hatte Richard Thaler, der damals an dem sehr konservativen Fachbereich Wirtschaftswissenschaften der Universität Rochester studierte, erstmals häretische Gedanken. Thaler hatte von jeher einen scharfen Verstand und eine Neigung zur Ironie, und als Student amüsierte er sich damit, Verhaltensbeobachtungen zusammenzutragen, die das Modell vom rationalen ökonomischen Verhalten nicht erklären konnte. Mit besonderer Wonne registrierte er Hinweise auf ökonomische Irrationalität bei seinen Professoren, und eines dieser Beispiele war besonders frappierend.

Professor R. (der mittlerweile als Richard Rosett identifiziert und später Dekan der Betriebswirtschaftlichen Fakultät der Universität Chicago wurde) war ein entschiedener Anhänger der ökonomischen Standardtheorie sowie ein anspruchsvoller Weinliebhaber. Thaler fiel auf, dass Professor R. nur äußerst ungern eine Flasche aus seiner Sammlung verkaufte – auch nicht für einen hohen Preis von 100 Dollar (in Dollar von 1975!). Professor R. kaufte Wein auf Auktionen, aber er zahlte nie mehr als 35 Dollar für eine Flasche dieser Qualität. Bei Preisen zwischen 35 und 100 Dollar kaufte und verkaufte er nicht. Die große Lücke ist inkonsistent mit der ökonomischen Theorie, der zufolge die Flasche für den Professor einen genau definierten Wert haben sollte. Wenn ihm eine bestimmte Flasche 50 Dollar wert sei, dann sollte er bereit sein, sie für jeden Betrag über 50 Dollar zu verkaufen. Wenn ihm die Flasche nicht gehörte, sollte er bereit sein, jeden Betrag bis zur Höhe von 50 Dollar dafür zu bezahlen. [3] Der gerade noch akzeptable Verkaufspreis und der gerade noch akzeptable Kaufpreis hätten identisch sein sollen, tatsächlich aber war der minimale Verkaufspreis (100 Dollar) viel höher als der maximale Kaufpreis von 35 Dollar. Der Besitz des Guts schien seinen Wert zu steigern.

Richard Thaler fand viele Beispiele für das, was er den »Endowment-Effekt« (»Besitztumseffekt«) nannte, insbesondere für Güter, die üblicherweise nicht gehandelt werden. Sie können sich leicht in eine

ähnliche Situation hineindenken. Angenommen, Sie besitzen eine Eintrittskarte für ein ausverkauftes Konzert einer populären Band, die sie zum regulären Preis von 200 Dollar gekauft haben. Sie sind ein absoluter Fan, und Sie wären bereit gewesen, bis zu 500 Dollar für die Eintrittskarte zu bezahlen. Jetzt haben Sie das Ticket und erfahren im Internet, dass reichere oder noch größere Fans 3000 Dollar bieten. Würden Sie verkaufen? Wenn Sie den meisten Besuchern ausverkaufter Veranstaltungen gleichen, verkaufen Sie nicht. Ihr niedrigster Verkaufspreis liegt über 3000 Dollar, und Ihr maximaler Kaufpreis beträgt 500 Dollar. Dies ist ein Beispiel für einen Endowment-Effekt, und jemand, der fest von der ökonomischen Standardtheorie überzeugt ist, wäre darüber verwundert.[4] Thaler suchte nach einer Erklärung für solche der herrschenden Theorie widersprechenden Beispiele.

Zufälligerweise machte Thaler bei einer Konferenz die Bekanntschaft eines unserer ehemaligen Studenten, der ihm einen frühen Entwurf der Neuen Erwartungstheorie überreichte. Er berichtet, er habe das Manuskript mit großer Begeisterung gelesen, weil ihm rasch klar wurde, dass die verlustaverse Wertfunktion der Neuen Erwartungstheorie den Endowment-Effekt und einige andere ungelöste Probleme in seiner Sammlung erklären könne. Die Lösung bestand darin, die Standardannahme, der Zustand, eine bestimmte Flasche zu *besitzen*, habe für Professor R. einen einzigartigen Nutzen, aufzugeben. Die Neue Erwartungstheorie behauptete, die Bereitschaft, die Flasche zu kaufen oder zu verkaufen, hänge von dem Referenzpunkt ab – ob der Professor die Flasche zum jetzigen Zeitpunkt besitzt oder nicht. Wenn er sie besitzt, erwägt er das Unlustgefühl, das mit dem *Weggeben* der Flasche verbunden ist. Wenn er sie nicht besitzt, erwägt er die Lust, die mit ihrem *Erwerb* verbunden ist. Aufgrund der Verlustaversion waren die Werte unterschiedlich: Das Weggeben einer Flasche guten Weins erzeugt mehr Unlust, als der Erwerb einer ebenso guten Flasche Lust erzeugt.[5] Erinnern Sie sich an die grafische Darstellung von Verlusten und Gewinnen im vorangehenden Kapitel. Die Steigung der Funktion ist im negativen Bereich steiler; die Reaktion auf einen Verlust ist stärker als die Reaktion auf einen entsprechenden Gewinn. Dies war die Erklärung des Endowment-Effekts, nach der Thaler gesucht hatte. Und die erste Anwendung der Neuen Erwartungstheorie auf ein schwieriges

ökonomisches Problem erscheint heute als ein bedeutender Meilenstein in der Entwicklung der Verhaltensökonomik.

Thaler arrangierte einen einjährigen Forschungsaufenthalt an der Universität Stanford, als er wusste, dass Amos und ich dort seien. Während dieser produktiven Zeit lernten wir viel voneinander und wurden Freunde. Sieben Jahre später hatten er und ich eine weitere Gelegenheit, ein Jahr gemeinsam zu verbringen und unser Gespräch zwischen Psychologie und Wirtschaftswissenschaften fortzusetzen. Die Russell Sage Foundation, die lange Zeit der wichtigste Finanzier verhaltensökonomischer Forschungen war, gewährte Thaler eines ihrer ersten Stipendien zu dem Zweck, ein Jahr mit mir in Vancouver zu forschen. Während dieses Jahres arbeiteten wir eng mit Jack Knetsch zusammen, einem an der dortigen Universität lehrenden Ökonomen, der sich wie wir intensiv für den Endowment-Effekt, die Regeln ökonomischer Fairness und scharfes chinesisches Essen interessierte.

Der Ausgangspunkt unserer Forschungsarbeiten war die Feststellung, dass der Endowment-Effekt nicht universell ist. Wenn jemand Sie bittet, einen Fünf-Dollar-Schein gegen fünf Ein-Dollar-Scheine zu tauschen, tun Sie dies, ohne im Geringsten das Gefühl eines Verlustes zu haben. Und auch wenn Sie Schuhe einkaufen gehen, spielt die Verlustaversion keine große Rolle. Der Händler, der die Schuhe gegen Geld weggibt, hat bestimmt nicht das Gefühl, etwas zu verlieren. Tatsächlich waren die Schuhe, die er übereignet, aus seiner Sicht von jeher nur ein platzraubender Ersatz *(proxy)* für jenes Geld, das er von einem Käufer einzusammeln hoffte. Überdies haben Sie höchstwahrscheinlich das Bezahlen des Händlers nicht als Verlust erlebt, weil Sie das Geld faktisch als einen Ersatz für die Schuhe zurückhielten, die Sie kaufen wollten. Diese Fälle von routinemäßigen Handelsgeschäften unterscheiden sich im Grunde nicht vom Wechseln eines Fünf-Dollar-Scheins in fünf Ein-Dollar-Scheine. Bei routinemäßigen Handelsgeschäften spielt die Verlustaversion auf beiden Seiten keine Rolle.

Was unterscheidet diese Markttransaktionen von dem Widerwillen des Professor R., seinen Wein zu verkaufen, oder von dem Widerwillen der Besitzer von Super-Bowl-Tickets, ihre Karten selbst zu einem sehr hohen Preis zu verkaufen? [6] Das unterscheidende Merkmal besteht darin, dass sowohl die Schuhe, die Ihnen der Händler verkauft, als auch das Geld, das Sie von ihrem Schuh-Budget ausgeben, »zum Tausch« gehalten werden. Sie sollen gegen andere Güter eingetauscht werden. Andere Güter wie Wein

und Super-Bowl-Tickets werden »zur Nutzung« bereitgehalten, also zum Verbrauch oder für eine andere genussvolle Verwendung. Unsere Freizeit und unser einkommensabhängiger Lebensstandard stehen ebenfalls nicht zum Verkauf oder zum Tausch.

Knetsch, Thaler und ich entwarfen ein Experiment, das den Gegensatz zwischen Gütern, die zum Verbrauch gehalten werden, und solchen, die zum Tausch bestimmt sind, unterstreichen sollte. Einen Aspekt des Versuchsdesigns übernahmen wir von Vernon Smith, dem Begründer der experimentellen Wirtschaftsforschung, mit dem ich viele Jahre später den Nobelpreis teilen sollte. Bei dieser Methode wird eine begrenzte Anzahl von Gutscheinen an die Teilnehmer in einem »Markt« verteilt. Alle Teilnehmer, die am Ende des Experiments einen Gutschein besitzen, können diesen gegen Bargeld einlösen. Der Rückzahlungswert ist für verschiedene Individuen unterschiedlich hoch, um die Tatsache zu berücksichtigen, dass die auf Märkten gehandelten Güter für einige Menschen wertvoller sind als für andere. Derselbe Gutschein mag für Sie 10 Dollar und für mich 20 Dollar wert sein, und ein Tauschgeschäft zu jedem beliebigen Preis zwischen diesen Werten wird für uns beide vorteilhaft sein.

Smith veranschaulichte mit seinen Experimenten auf plastische Weise, wie gut die grundlegenden Mechanismen von Angebot und Nachfrage funktionieren. Die Probanden machten aufeinanderfolgende öffentliche Angebote zum Kauf oder Verkauf eines Gutscheins, und andere reagierten öffentlich auf das Angebot. Jeder verfolgt diese Tauschgeschäfte und sieht den Preis, zu dem die Gutscheine den Besitzer wechseln. Die Resultate sind so regelmäßig wie die eines physikalischen Versuchs. So sicher, wie Wasser bergab fließt, verkaufen diejenigen, die einen Gutschein besitzen, der für sie von geringem Wert ist (weil sein Rückzahlungswert niedrig ist), schließlich ihren Gutschein mit Gewinn an jemanden, der ihm einen höheren Wert zuschreibt. Wenn die Handelsgeschäfte aufhören, befinden sich die Gutscheine in den Händen derjenigen, die von dem Experimentator am meisten Geld dafür bekommen können. Die Magie des Marktes hat funktioniert! Außerdem sagt die volkswirtschaftliche Theorie sowohl den Endpreis, auf den sich der Markt einigen wird, als auch die Anzahl der Gutscheine, die den Besitzer wechseln werden, zutreffend vorher. Wenn der Hälfte der Marktteilnehmer nach dem Zufallsprinzip Gutscheine zugeteilt

werden, dann sagt die Theorie vorher, dass die Hälfte der Gutscheine ihren Besitzer wechseln wird.[7]

Wir benutzten für unser Experiment eine Variante von Smiths Methode. Jede Sitzung begann mit mehreren Gutscheinhandelsrunden, die Smiths Ergebnisse exakt reproduzierten. Die geschätzte Anzahl von Handelsgeschäften lag in der Regel sehr nahe bei der von der Standardtheorie vorhergesagten Anzahl oder war sogar mit dieser identisch. Die Gutscheine besaßen selbstverständlich nur deshalb einen Wert, weil sie gegen das Bargeld des Experimentators eingetauscht werden konnten; sie hatten keinen Gebrauchswert. Anschließend konzipierten wir einen ähnlichen Markt für ein Objekt, von dem wir annahmen, dass es für die Teilnehmer einen Gebrauchswert besitzen würde: eine attraktive Kaffeetasse, die mit den Emblemen der jeweiligen Universität verziert war, an der wir die Experimente durchführten. Die Kaffeetasse war damals etwa 6 Dollar wert (und sie wäre heute etwa das Doppelte wert). Die Tassen wurden nach dem Zufallsprinzip an die Hälfte der Teilnehmer verteilt. Die Verkäufer hatten ihre Tasse vor sich stehen, die Käufer sollten sich die Tasse ihres Nachbarn ansehen; alle gaben die Preise an, zu denen sie handeln würden. Die Käufer mussten eine Tasse mit ihrem eigenen Geld erwerben. Die Ergebnisse waren dramatisch: Der durchschnittliche Verkaufspreis betrug etwa das Doppelte des durchschnittlichen Kaufpreises, und die geschätzte Anzahl der Handelsgeschäfte betrug weniger als die Hälfte der Anzahl, die von der Standardtheorie vorhergesagt wurde. Die Magie des Marktes funktionierte nicht für ein Gut, das für die Besitzer einen Gebrauchswert besaß.

Wir führten eine Reihe von Experimenten durch, bei denen wir Varianten der gleichen Vorgehensweise benutzten – immer mit denselben Ergebnissen. Mein Lieblingsexperiment ist eines, bei dem wir den Verkäufern und Käufern eine dritte Gruppe zugesellten – die Wähler. Anders als die Käufer, die ihr eigenes Geld ausgeben mussten, um das Gut zu erwerben, konnten die Wähler entweder eine Tasse oder eine Summe Geld erhalten, und sie gaben den Geldbetrag an, der genauso begehrenswert war wie das Gut. Dies waren die Ergebnisse:

| | |
|---|---|
| Verkäufer | 7,12 Dollar |
| Wähler | 3,12 Dollar |
| Käufer | 2,87 Dollar |

Die Diskrepanz zwischen Verkäufern und Wählern ist bemerkenswert, weil sie tatsächlich vor der gleichen Wahl stehen! Ein Verkäufer kann entweder mit einer Tasse oder mit Geld nach Hause gehen, ein Wähler hat genau die gleichen beiden Optionen. Die langfristigen Folgen der Entscheidung sind für beide Gruppen gleich. Der einzige Unterschied liegt in der Emotion des Augenblicks. Der hohe Preis, den Verkäufer festsetzen, spiegelt die Tatsache wider, dass sie ein Objekt, welches sie bereits besitzen, nur widerstrebend abgeben – ein Widerstreben, das man bereits bei Säuglingen beobachten kann, die sich fest an ein Spielzeug klammern und eine starke motorische Unruhe zeigen, wenn es ihnen weggenommen wird. Die Verlustaversion ist in die automatischen Bewertungen von System 1 eingebaut.

Käufer und Wähler setzen ähnliche Barwerte fest, obwohl die Käufer für die Tasse zahlen müssen, die für die Wähler gratis ist. Dies sollten wir erwarten, wenn die Käufer das Ausgeben von Geld für die Tasse nicht als einen Verlust erleben. Neuroimaging-Studien bestätigen den Unterschied. Der Verkauf von Gütern, die man normalerweise benutzen würde, aktiviert Hirnregionen, die mit Ekel und Schmerz assoziiert sind. Auch das Kaufen von Gütern aktiviert diese Areale, aber nur, wenn die Preise als zu hoch wahrgenommen werden – wenn man das Gefühl hat, dass ein Verkäufer mehr Geld nimmt, als dem Tauschwert entspricht. Hirnscan-Untersuchungen deuten auch darauf hin, dass Einkäufe zu sehr niedrigen Preisen eine lustvolle Erfahrung sind.[8]

Der Barwert, den die Verkäufer der Tasse zuschreiben, ist etwas mehr als doppelt so hoch wie der Wert, den Wähler und Käufer für sie festsetzen. Der Quotient liegt sehr nahe an dem Verlustaversionskoeffizienten bei riskanten Wahlen, wie zu erwarten ist, wenn wir sowohl auf risikolose als

auch auf riskante Entscheidungen die gleiche Wertfunktion für Geldgewinne und -verluste anwenden.[9] In zahlreichen Studien über diverse ökonomische Sachverhalte einschließlich der Reaktion von Haushalten auf Preisänderungen hat sich ein Verhältnis von etwa 2:1 herauskristallisiert. Entsprechend den Vorhersagen von Ökonomen kaufen Verbraucher bei sinkenden Preisen tendenziell mehr Eier, Orangensaft oder Fisch, während sie bei steigenden Preisen weniger kaufen; doch im Gegensatz zu den Vorhersagen der ökonomischen Theorie ist der Effekt von Preissteigerungen (Verlusten relativ zum Referenzpreis) etwa doppelt so groß wie der Effekt von Preissenkungen (Gewinnen).[10]

Das Tassen-Experiment ist noch immer die Standard-Demonstration des Endowment-Effekts, zusammen mit einem noch einfacheren Experiment, über das Jack Knetsch ungefähr zur gleichen Zeit berichtete. Knetsch bat zwei Klassen, einen Fragebogen auszufüllen, und er belohnte die Teilnehmer mit einem Geschenk, das für die Dauer des Experiments sichtbar vor ihnen lag. In einer Sitzung war die Belohnung ein teurer Füller, in der anderen ein Riegel Schweizer Schokolade. Am Ende der Sitzung zeigte der Experimentator das alternative Geschenk und erlaubte allen, ihr Geschenk gegen ein anderes einzutauschen. Nur etwa 10 Prozent der Teilnehmer entschieden sich dafür, ihr Geschenk zu tauschen. Die meisten derjenigen, die den Füller bekommen hatten, blieben bei dem Füller, und diejenigen, die die Schokolade bekommen hatten, blieben ebenfalls dabei.

## Wie ein Wertpapierhändler denken

Die grundlegenden Annahmen der Neuen Erwartungstheorie lauten, dass es Referenzpunkte gibt und dass Verluste stärker ins Gewicht fallen als entsprechende Gewinne. Daten über reale Märkte, die im Lauf der Jahre zusammengetragen wurden, verdeutlichen die Erklärungskraft dieser Konzepte.[11] Eine Studie über den Markt für Eigentumswohnungen in Boston während einer Rezession ergab besonders klare Ergebnisse.[12] Die Autoren dieser Studie verglichen das Verhalten von Eigentümern ähnlicher Wohnungen, die diese zu unterschiedlichen Preisen gekauft hatten. Für einen rationalen Agenten ist der Kaufpreis eine irrelevante Tatsache der Vergangenheit – allein der gegenwärtige Marktwert zählt. Nicht so für *Humans* in einem Immobilienmarkt mit rückläufigen Preisen. Eigentümer, die einen hohen Referenzpunkt haben und damit höhere Verluste riskieren, schreiben ihrer Wohnung einen höheren Wert zu, versuchen länger, ihre Wohnung zu verkaufen, und bekommen schließlich mehr Geld dafür.

Der ursprüngliche Nachweis einer Asymmetrie zwischen Verkaufs- und Kaufpreisen (beziehungsweise, noch überzeugender, zwischen Verkaufen und Auswählen) war sehr wichtig für die anfängliche Akzeptanz der Konzepte von Referenzpunkt und Verlustaversion. Es besteht jedoch Einvernehmen darüber, dass Referenzpunkte instabil sind, insbesondere in ungewöhnlichen Laborsituationen, und dass der Endowment-Effekt dadurch beseitigt werden kann, dass man den Referenzpunkt verändert.

Kein Endowment-Effekt wird erwartet, wenn Eigentümer ihre Güter als »Wertträger« für zukünftige Tauschgeschäfte betrachten, eine weitverbreitete Einstellung auf den üblichen Handels- und Finanzmärkten. Der Experimentalökonom John List, der das Handeln bei Baseballkarten-Tauschbören erforscht hat, stellte fest, dass Neulinge in diesem Geschäft sich nur widerwillig von ihren Karten trennten, aber dass dieser Widerwille schließlich mit zunehmender Handelserfahrung verschwand. Noch überraschender war, dass List einen starken Effekt der Handelserfahrung auf den Endowment-Effekt für neue Güter feststellte.[13]

Bei einer Tauschbörse stellte List ein Schild auf, das Passanten bat, an einer kurzen Umfrage teilzunehmen, für die sie mit einem kleinen

Geschenk belohnt würden: einer Kaffeetasse oder einer gleichwertigen Tafel Schokolade. Die Geschenke wurden nach dem Zufallsprinzip verteilt. Als die Teilnehmer aufbrechen wollten, sagte List zu jedem von ihnen: »Wir haben Ihnen eine Tasse [oder eine Tafel Schokolade] gegeben, aber Sie können sie gegen eine Tafel Schokolade [oder eine Tasse] eintauschen, wenn Sie wollen.« Lists Ergebnisse stimmten exakt mit den Befunden von Jack Knetschs früheren Experimenten überein: Er stellte fest, dass nur 18 Prozent der unerfahrenen Händler bereit waren, ihr Geschenk gegen das andere einzutauschen. In deutlichem Gegensatz dazu zeigten erfahrene Händler keine Spur von einem Endowment-Effekt: 48 Prozent von ihnen tauschten! Zumindest in einem Marktumfeld, in dem Handeln die Regel war, waren sie Tauschgeschäften nicht abgeneigt.

Jack Knetsch führte auch Experimente durch, bei denen subtile Manipulationen den Endowment-Effekt zum Verschwinden brachten.[14] Teilnehmer zeigten nur dann einen Endowment-Effekt, wenn sie das Gut eine Zeit lang physisch besaßen, ehe die Möglichkeit, damit zu handeln, erwähnt wurde. Verfechter der ökonomischen Standardtheorie würden vielleicht ins Feld führen, Knetsch habe zu viel Zeit mit Psychologen verbracht, weil seine experimentelle Manipulation ein Interesse an den Variablen verrate, die Sozialpsychologen für wichtig hielten. Tatsächlich sind in der laufenden Debatte über den Endowment-Effekt deutliche Unterschiede in den methodischen Ansätzen von Experimentalökonomen und -psychologen zutage getreten.[15]

Erfahrene Händler haben offenkundig gelernt, die richtige Frage zu stellen, die lautet: »Wie sehr will ich diese Tasse *besitzen*, im Vergleich zu anderen Dingen, die ich stattdessen haben könnte?« Dies ist die Frage, die *Econs* stellen, und bei dieser Frage gibt es keinen Endowment-Effekt, weil die Asymmetrie zwischen der Lust, etwas zu bekommen, und der Unlust, etwas wegzugeben, hier irrelevant ist.

Neueste Studien über die Psychologie der »Entscheidungsfindung unter Armut« deuten darauf hin, dass die Armen eine weitere Gruppe sind, bei der wir ebenfalls keinen Endowment-Effekt erwarten. In der Neuen Erwartungstheorie bedeutet Armut, unterhalb des eigenen Referenzpunktes zu leben. Es gibt Güter, die die Armen benötigen und sich nicht leisten können, sodass sie immer »in der Verlustzone« sind. Kleine Geldbeträge, die sie erhalten, werden daher als ein verminderter Verlust, nicht als

Gewinn wahrgenommen. Das Geld hilft einem, ein wenig in Richtung des Referenzpunktes zu klettern, aber die Armen bleiben immer auf dem steilen Schenkel der Wertfunktion.

Menschen, die arm sind, denken wie Händler, aber die Dynamik ist eine ganz andere.[16] Anders als Händler sind die Armen nicht indifferent gegenüber den Unterschieden zwischen Gewinnen und Weggeben. Ihr Problem ist, dass sie nur zwischen Verlusten wählen können. Geld, das sie für ein Gut ausgeben, bedeutet den Verlust eines anderen Gutes, das sie stattdessen hätten kaufen können. Für die Armen sind Kosten Verluste.

Wir alle kennen Menschen, denen es wehtut, Geld auszugeben, obwohl sie, objektiv gesehen, recht wohlhabend sind. Es mag auch kulturelle Unterschiede in der Einstellung zu Geld geben und insbesondere zu Ausgaben für Impulskäufe und kleinere Luxusartikel, wie etwa den Kauf einer verzierten Tasse. Ein solcher Unterschied mag die große Diskrepanz zwischen den Ergebnissen der »Tassen-Studie« in den Vereinigten Staaten und in Großbritannien erklären.[17] Kauf- und Verkaufspreise unterscheiden sich erheblich bei Experimenten, die mit zufällig ausgewählten Studenten in den USA durchgeführt wurden, aber die Unterschiede sind viel geringer als bei englischen Studenten. Es gibt noch viel über den Endowment-Effekt zu lernen.

### Zum Thema »Endowment-Effekt«

> »Ihr war egal, welches der beiden Büros sie bekommen würde, aber einen Tag nach der Bekanntgabe war sie nicht länger gewillt, zu tauschen. Endowment-Effekt !«

> »Diese Verhandlungen führen nirgendwohin, weil es beiden Seiten schwerfällt, Zugeständnisse zu machen, selbst wenn sie im Gegenzug etwas dafür bekommen. Verluste werden höher bewertet als Gewinne.«

»Als sie die Preise anhoben, versiegte die Nachfrage.«

»Er hasst den Gedanken, sein Haus für weniger Geld zu verkaufen, als er dafür gezahlt hatte. Hier wirkt die Verlustaversion.«

»Er ist ein Geizhals und empfindet jeden Dollar, den er ausgibt, als einen Verlust.«

# *28. Negative Ereignisse*

Das Konzept der Verlustaversion ist zweifellos der wichtigste Beitrag der Psychologie zur Verhaltensökonomik. Dies ist seltsam, weil es niemanden überrascht, dass Menschen viele Ergebnisse als Gewinne und Verluste beurteilen und Verluste höher bewerten als Gewinne. Amos und ich sagten oft im Scherz, wir erforschten ein Gebiet, über das unsere Großmütter eine Menge wüssten. Tatsächlich aber wissen wir mehr als unsere Großmütter, und wir können die Verlustaversion jetzt in den Rahmen eines umfassenderen Zwei-Systeme-Modells des Geistes integrieren, insbesondere in eine biologische und psychologische Betrachtungsweise, in der Negativität und Flucht stärker sind als Positivität und Annäherung. Die Folgen der Verlustaversion lassen sich in erstaunlich unterschiedlichen Phänomenen aufspüren: nur Mehraufwendungen werden ersetzt, wenn Güter beim Transport verloren gehen; Versuche zu umfassenden Reformen scheitern sehr oft; Profigolfer putten treffsicherer zum Par als zum Birdie. Klug, wie sie war, wäre meine Großmutter erstaunt gewesen über die konkreten Vorhersagen aus einer allgemeinen Idee, die sie für selbstverständlich hielt.

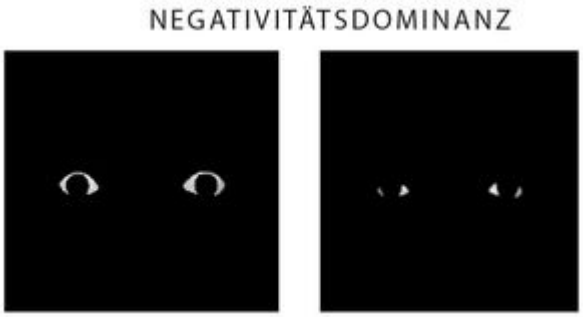


**Abbildung 12**

Ihr Herz schlug schneller, als Sie das linke Bild betrachteten.[1] Es schlug schon schneller, bevor sie das, was an dem Bild so unheimlich war, benennen konnten. Nach einiger Zeit erkannten Sie vielleicht die Augen einer erschrockenen Person. Die Augen rechts, die durch die gehobenen

Wangen eines Lächelns verengt sind, drücken Freude aus – und sie rufen keine annähernd so starke Erregung hervor. Die beiden Bilder wurden Menschen gezeigt, die in einem Hirnscanner lagen. Jedes Bild wurde kürzer als eine Zweihundertstelsekunde dargeboten und sofort durch »visuelles Rauschen« maskiert, eine zufällige Darbietung dunkler und heller Quadrate. Keiner der Betrachter hatte bewusst registriert, dass er Bilder von Augen gesehen hatte, aber ein Teil ihres Gehirns »wusste« es ganz offensichtlich: die Amygdala, die vor allem als »Bedrohungszentrum« des Gehirns fungiert, obgleich sie auch bei anderen emotionalen Zuständen aktiviert ist. Die Hirnaufnahmen zeigten eine starke Reaktion der Amygdala auf ein bedrohliches Bild, das der Betrachter nicht als solches erkannte. Die Information über die Bedrohung wurde wahrscheinlich über einen superschnellen Nervenkanal weitergeleitet, der direkt in ein Hirnareal mündet, das Emotionen verarbeitet, und dabei die Sehrinde umgeht, in der das bewusste Erleben des »Sehens« stattfindet.[2] Derselbe neuronale Schaltkreis ist auch dafür verantwortlich, dass schematische wütende Gesichter (eine potenzielle Bedrohung) schneller und effizienter verarbeitet werden als schematische zufriedene Gesichter. [3] Einige Experimentatoren haben berichtet, dass ein wütendes Gesicht aus einer Menge von heiteren Gesichtern heraussticht, während ein einzelnes heiteres Gesicht sich nicht aus einer wütenden Menge heraushebt.[4] Das Gehirn des Menschen und anderer Säugetiere enthält einen Mechanismus, der darauf ausgelegt ist, schlechten Nachrichten den Vorrang zu geben. Indem dieser Schaltkreis die Zeitspanne für die Wahrnehmung eines Fressfeindes um ein paar Hundertstelsekunden verkürzt, erhöht er die Chancen des Tieres, lange genug zu leben, um sich fortzupflanzen. In den automatischen Operationen von System 1 spiegelt sich diese evolutionäre Geschichte wider. Bislang wurde kein vergleichbar schneller Mechanismus für das Erkennen guter Neuigkeiten entdeckt. Natürlich reagieren wir und unsere tierischen Verwandten schnell auf Anzeichen für Paarungs- und Ernährungschancen, und Werbeagenturen gestalten die Plakatflächen entsprechend. Trotzdem werden Bedrohungen gegenüber Chancen privilegiert, wie es im Interesse der Selbsterhaltung auch sein sollte.

Selbst auf rein symbolische Bedrohungen reagiert das Gehirn schnell. Emotional aufgeladene Wörter ziehen rasch die Aufmerksamkeit auf sich, und negativ besetzte Wörter (»Krieg«, »Verbrechen«) ziehen die

Aufmerksamkeit schneller auf sich als positiv besetzte (»Frieden«, »Liebe«). Es gibt keine reale Bedrohung, aber die bloße Erinnerung an ein negatives Ereignis wird von System 1 als bedrohlich eingestuft. Wie wir früher bei dem Wort »erbrechen« sahen, löst die symbolische Repräsentation in abgeschwächter Form assoziativ viele der Reaktionen aus, die der reale Sachverhalt hervorruft, einschließlich physiologischer Anzeichen emotionaler Zustände und ansatzweiser Tendenzen zur Vermeidung oder Annäherung, zum Zurückschrecken oder Vorneigen. Die Empfindlichkeit für Bedrohungen erstreckt sich auch auf die Verarbeitung von Meinungsäußerungen, mit denen wir nicht übereinstimmen. Je nach Ihrer Einstellung zur Euthanasie bräuchte Ihr Gehirn weniger als eine Viertelsekunde, um die »Bedrohung« in einem Satz zu bemerken, der so beginnt: »Ich halte Euthanasie für eine akzeptable/inakzeptable …«[5]

Der Psychologe Paul Rozin, ein Experte auf dem Gebiet der Ekel-Forschung, hat beobachtet, dass eine einzige Schabe die Anziehungskraft einer Schüssel Kirschen völlig ruinieren würde, während eine Kirsche in einer Schüssel voller Schaben keinerlei Unterschied macht. Er weist darauf hin, dass das Negative in vielfacher Weise das Positive aussticht, und die Verlustaversion ist eine von vielen Manifestationen einer allgemeinen Negativitätsdominanz.[6] Andere Wissenschaftler fassten den Forschungsstand in einem Aufsatz mit dem Titel »Bad Is Stronger Than Good« (»Schlecht ist stärker als gut«) folgendermaßen zusammen: »Negative Emotionen, schlechte Eltern und schlechtes Feedback haben eine stärkere Wirkung als ihre positiven Pendants, und negative Informationen werden sorgfältiger verarbeitet als positive. Das Selbst ist stärker motiviert, negative Selbstdefinitionen zu vermeiden, als positive zu verfolgen. Ein negativer Eindruck und negative Stereotype entstehen schneller und sind widerlegungsresistenter als ihre positiven Entsprechungen.«[7] Sie zitieren John Gottman, den bekannten Experten für eheliche Beziehungen, der feststellte, dass das langfristige Gelingen einer Beziehung viel stärker von der Vermeidung des Negativen als von dem Streben nach dem Positiven abhängig ist. Nach Gottmans Überzeugung erfordert eine stabile Beziehung, dass positive Interaktionen die negativen mindestens im Verhältnis von 5:1 übertreffen. Andere Asymmetrien im sozialen Bereich sind noch frappierender. Wir alle wissen, dass eine Freundschaft, die

vielleicht Jahre brauchte, um sich zu entwickeln, durch eine einzige Handlung ruiniert werden kann.

Einige Unterschiede zwischen Gut und Schlecht sind biologisch fest verdrahtet. Schon Neugeborene empfinden offenkundig Schmerzen als etwas Negatives und Süßes (bis zu einem gewissen Grad) als etwas Positives. Doch in vielen Situationen ist die Grenze zwischen Gut und Schlecht ein Referenzpunkt, der sich mit der Zeit verändert und von den unmittelbaren Umständen abhängig ist. Stellen Sie sich vor, Sie wären in einer kalten Nacht draußen unterwegs, unzureichend geschützt gegen die sintflutartigen Regenfälle, Ihre Kleidung klatschnass. Ein beißend kalter Wind macht Ihr Elend komplett. Herumirrend finden Sie einen großen Felsen, der einen gewissen Schutz vor den entfesselten Elementen bietet. Der Biologe Michel Cabanac würde die Erfahrung dieses Moments als höchst lustvoll beschreiben, weil sie, wie es Lust normalerweise tut, die Richtung einer biologisch bedeutsamen Verbesserung der Situation anzeigt.[8] Die lustvolle Erleichterung wird natürlich nicht sehr lange dauern, und Sie werden bald auch hinter dem Felsen wieder schlottern und, getrieben von Ihrer erneuten Pein, nach einem besseren Unterschlupf suchen.

## Ziele sind Referenzpunkte

Die Verlustaversion bezieht sich auf die relative Stärke zweier Motive: Wir sind stärker motiviert, Verluste zu vermeiden, als Gewinne zu erzielen. Ein Referenzpunkt ist manchmal der Status quo, aber er kann auch ein Ziel in der Zukunft sein: Ein Ziel nicht zu erreichen ist ein Verlust, das Ziel zu übertreffen ein Gewinn. Wie wir aus der Negativitätsdominanz erwarten würden, sind die beiden Motive nicht gleich stark.[9] Die Aversion gegen das Nichterreichen des Ziels ist viel stärker als das Verlangen, es zu übertreffen.

Menschen setzen sich oft kurzfristige Ziele, die sie erreichen, aber nicht unbedingt übertreffen wollen. Sie werden in ihrer Anstrengung vermutlich nachlassen, wenn sie ein Zwischenziel erreicht haben, mit Ergebnissen, die manchmal der ökonomischen Logik widersprechen. New Yorker Taxifahrer zum Beispiel mögen ein monatliches oder jährliches Zieleinkommen haben, aber das Ziel, das ihren Leistungsantrieb bestimmt, ist in der Regel ein tägliches Verdienstziel. Das tägliche Ziel ist selbstverständlich an manchen Tagen viel leichter zu erreichen (und zu übertreffen) als an anderen. An regnerischen Tagen bleibt ein New Yorker Taxi nie lange frei, und der Fahrer erreicht sein Ziel sehr schnell; nicht so bei gutem Wetter, wenn Taxis oft Zeit damit verschwenden, auf der Suche nach Fahrgästen Straßen abzufahren. Aus der ökonomischen Logik folgt, dass Taxifahrer an regnerischen Tagen Überstunden machen sollten, während sie sich an milden Tagen ein bisschen Freizeit genehmigen sollten, denn dann können sie Freizeit zu einem niedrigeren Preis »kaufen«. Die Logik der Verlustaversion legt das Gegenteil nahe: Fahrer, die ein festes Tagesziel haben, werden viele Überstunden machen, wenn sie wenig Fahrgäste einsammeln, und sie werden früh Feierabend machen, wenn viele regendurchnässte Kunden am Straßenrand auf ein Taxi warten.[10]

Die Ökonomen Devin Pope und Maurice Schweitzer von der Universität von Pennsylvania gelangten zu der Überzeugung, dass Golf ein perfektes Beispiel für einen Referenzpunkt liefert: das Par. Mit jedem Loch auf dem Golfplatz ist eine bestimmte Anzahl an Schlägen verbunden; die Par-Zahl stellt die Basislinie für eine gute – wenn auch keine überragende – Leistung dar. Für einen Profigolfer ist ein Birdie (ein Schlag unter Par) ein Gewinn

und ein Bogey (ein Schlag über Par) ein Verlust. Die Ökonomen verglichen zwei Situationen, in denen sich ein Spieler befinden kann, wenn er in der Nähe des Lochs ist:

– Putten, um einen Bogey zu vermeiden
– Putten, um einen Birdie zu erzielen

Beim Golfen zählt jeder Schlag, und beim Profigolf zählt jeder Schlag eine Menge. Doch laut der Neuen Erwartungstheorie zählen einige Schläge mehr als andere. Das Par nicht zu erreichen ist ein Verlust, aber einen Birdie-Putt zu verfehlen ist ein entgangener Gewinn, kein Verlust. Pope und Schweitzer folgerten aus der Verlustaversion, dass die Spieler sich ein wenig stärker anstrengen würden, wenn sie für ein Par putten (um einen Bogey zu vermeiden), als wenn sie für einen Birdie putten. Sie analysierten sehr detailliert über 2,5 Millionen Putts, um diese Vorhersage zu überprüfen.

Sie hatten recht. Unabhängig davon, ob der Putt leicht oder schwer war, waren die Spieler in jeder Entfernung vom Loch erfolgreicher, wenn sie für ein Par, nicht für einen Birdie putteten. Die Differenz in ihren Erfolgsquoten, je nachdem, ob sie es auf ein Par (um einen Bogey zu vermeiden) oder auf einen Birdie abgesehen hatten, betrug 3,6 Prozent. Diese Differenz ist nicht trivial. Tiger Woods war einer der »Teilnehmer« ihrer Studie. Wenn es Tiger Woods in seinen besten Jahren gelungen wäre, genauso gut für Birdies wie für Par zu putten, hätten sich sein durchschnittlicher Turnier-Score um einen Schlag und sein Verdienst um fast eine Million Dollar pro Saison verbessert. Diese Profis, die sich einem erbitterten Wettbewerb stellen müssen, treffen gewiss nicht die bewusste Entscheidung, bei Birdie-Putts nachzulassen, aber ihre starke Aversion gegen einen Bogey trägt scheinbar zu einer zusätzlichen Konzentration auf die anstehende Aufgabe bei.

Die Erforschung von Putts verdeutlicht, dass ein theoretisches Konzept als Denkhilfe sehr leistungsfähig sein kann. Wer hätte gedacht, dass es sich lohnen würde, monatelang Par- und Birdie-Putts zu analysieren? Das Konzept der Verlustaversion, das vielleicht bis auf einige Ökonomen niemanden überrascht, erzeugte eine präzise und nicht intuitive Hypothese und führte Forscher zu einer Entdeckung, die alle überraschte – Profigolfer eingeschlossen.

## Den Status quo verteidigen

Wenn man erst einmal danach Ausschau hält, findet man die asymmetrische Intensität der Motive zur Verlustvermeidung und Gewinnerzielung fast überall. Sie ist ein allgegenwärtiges Merkmal von Verhandlungen, insbesondere von Neuverhandlungen über einen bestehenden Vertrag, was die typische Situation bei Tarifverhandlungen und bei internationalen Handelsgesprächen oder Verhandlungen über Rüstungsbeschränkungen ist. Die bestehenden Regelungen definieren Referenzpunkte, und eine vorgeschlagene Änderung einer Klausel der Vereinbarung wird zwangsläufig als ein Zugeständnis gesehen, das eine Seite der anderen macht. Die Verlustaversion erzeugt eine Asymmetrie, die die Einigung erschwert. Die Zugeständnisse, die Sie mir machen, sind meine Gewinne, aber Ihre Verluste; sie verursachen Ihnen viel mehr Unlust, als sie mir Lust bescheren. Sie werden ihnen zwangsläufig einen höheren Stellenwert einräumen, als ich dies tue. Das Gleiche gilt natürlich für die sehr schmerzlichen Zugeständnisse, die Sie von mir verlangen und die Sie offenbar nicht hinreichend zu würdigen wissen! Verhandlungen über einen »schrumpfenden Kuchen« sind besonders schwierig, weil sie eine Zuteilung von Verlusten erfordern. Menschen sind im Allgemeinen viel unbeschwerter, wenn sie über einen »größer werdenden Kuchen« verhandeln.

Viele der Botschaften, die Unterhändler bei Verhandlungen austauschen, sind Versuche, einen Referenzpunkt zu kommunizieren und der anderen Seite einen Anker zu liefern.[11] Die Botschaften sind nicht immer aufrichtig. Unterhändler tun oftmals so, als würde ihnen ein Gut besonders viel bedeuten (vielleicht Raketen eines bestimmten Typs bei Abrüstungsverhandlungen), obgleich sie dieses Gut in Wirklichkeit als Unterpfand einsetzen und letztlich beabsichtigen, es im Gegenzug für etwas anderes aufzugeben. Weil Unterhändler von einer Norm der Gegenseitigkeit beeinflusst werden, verlangt ein als schmerzlich ausgegebenes Zugeständnis ein ebenso schmerzliches (und vielleicht in gleicher Weise bloß vorgetäuschtes) Zugeständnis der anderen Seite.

Säugetiere einschließlich des Menschen streben stärker danach, Verluste zu vermeiden, als Gewinne zu erzielen. In der Welt territorialer Tiere erklärt

dieses Prinzip den Erfolg der Revierverteidiger. Ein Biologe schrieb dazu: »Wenn ein Revierhalter von einem Rivalen herausgefordert wird, gewinnt der Halter fast immer den Kampf – in der Regel binnen Sekunden.«[12] In menschlichen Angelegenheiten erklärt die gleiche einfache Regel einen Großteil dessen, was geschieht, wenn sich Institutionen zu reformieren versuchen, etwa bei der »Neuordnung« und »Umstrukturierung« von Unternehmen und bei Bemühungen, einen bürokratischen Apparat zu straffen, das Steuergesetz zu vereinfachen oder die medizinischen Behandlungskosten zu senken. Reformpläne, so, wie sie ursprünglich ausgearbeitet wurden, produzieren fast immer viele Gewinner und einige Verlierer, während sie insgesamt eine Verbesserung erzielen. Wenn die von den Maßnahmen Betroffenen politischen Einfluss besitzen, werden die potenziellen Verlierer allerdings aktiver und entschlossener sein als die potenziellen Gewinner; das Ergebnis wird daher zu ihren Gunsten verzerrt und zwangsläufig kostspieliger und weniger effektiv als ursprünglich geplant sein. Reformen beinhalten normalerweise sogenannte Besitzstandsklauseln, die die gegenwärtigen Anspruchsberechtigten schützen – zum Beispiel, wenn die vorhandene Belegschaft durch »natürlichen« Arbeitskräfteabgang statt durch Entlassungen abgebaut wird oder wenn Lohnkürzungen und Streichungen bei den Zusatzleistungen nur für neu eingestellte Arbeitnehmer gelten. Die Verlustaversion ist eine starke konservative Kraft, die minimale Veränderungen des Status quo bei Institutionen und im Leben von Individuen begünstigt. Dieser Konservatismus erklärte unsere Tendenz, in unserem Wohnviertel, unserer Partnerschaft und an unserem Arbeitsplatz zu verharren; er ist die Schwerkraft, die unser Leben in der Nähe des Referenzpunkts zusammenhält.

## Das Rechtswesen und die Verlustaversion

In dem Jahr, in dem wir in Vancouver zusammenarbeiteten, wandten sich Richard Thaler, Jack Knetsch und ich der Erforschung der Fairness ökonomischer Transaktionen zu, zum Teil weil uns dieses Thema interessierte, aber auch weil wir eine Gelegenheit und eine Verpflichtung hatten, jede Woche einen neuen Fragebogen zu entwerfen. Das kanadische Fischereiministerium finanzierte ein Programm für arbeitslose Akademiker in Toronto, die dafür bezahlt wurden, telefonische Umfragen durchzuführen. Das große Team von Interviewern arbeitete jeden Abend, und ständig wurden neue Fragen benötigt, um die Operation am Laufen zu halten. Auf Vermittlung von Jack Knetsch erklärten wir uns bereit, jede Woche einen Fragebogen in vier farbcodierten Versionen abzuliefern. Wir konnten nach allem fragen; die einzige Einschränkung war, dass der Fragebogen mindestens eine Frage nach Fischen enthalten sollte, um einen sachlichen Bezug zum Auftrag des Ministeriums herzustellen. Dieses Projekt lief über viele Monate, und in dieser Zeit erhoben wir eine Unmenge von Daten.

Wir studierten öffentliche Wahrnehmungen dessen, was als unfaires Verhalten von Einzelhändlern, Arbeitgebern und Vermietern angesehen wurde.[13] Unsere übergeordnete Frage lautete, ob die mit unfairem Verhalten verbundene soziale Schmach dem Profitstreben gewisse Schranken auferlegt. Wir fanden heraus, dass dies der Fall ist. Wir fanden auch heraus, dass die moralischen Regeln, nach denen die Befragten beurteilten, was Firmen tun oder nicht tun sollten, einen ausschlaggebenden Unterschied zwischen Verlusten und Gewinnen machen. Das Grundprinzip lautet, dass der aktuelle Lohn, Preis oder Mietzins einen Referenzpunkt festsetzt, der seinem Wesen nach ein Anspruch ist, der nicht angetastet werden darf. Es gilt als unfair, wenn ein Unternehmen seinen Kunden oder Mitarbeitern in Bezug auf die Referenztransaktion Verluste aufbürdet, es sei denn, es ist zur Wahrung seiner eigenen Ansprüche dazu gezwungen. Betrachten Sie dieses Beispiel:

> Ein Baumarkt hat Schneeschaufeln für 15 Dollar verkauft. Am Morgen nach einem schweren Schneesturm erhöht der Baumarkt den Preis auf 20 Dollar.
>
> Bitte beurteilen Sie diesen Schritt als:
> völlig in Ordnung akzeptabel unfair sehr unfair

Der Baumarkt verhält sich nach dem ökonomischen Standardmodell völlig angemessen: Er reagiert auf die erhöhte Nachfrage mit einer Preiserhöhung. Die Teilnehmer der Umfrage waren anderer Ansicht: 82 Prozent beurteilten die Aktion als »unfair« oder »sehr unfair«. Offenkundig betrachteten sie den Preis vor dem Schneesturm als einen Referenzpunkt und den erhöhten Preis als einen Verlust, den der Baumarkt seinen Kunden auferlegte – nicht weil er dazu gezwungen war, sondern einfach weil er es konnte. Wir fanden heraus, dass eine grundlegende Fairnessregel besagt, dass es nicht akzeptabel ist, seine Marktmacht zu dem Zweck auszunutzen, anderen Verluste aufzuerlegen. Das folgende Beispiel veranschaulicht diese Regel in einem anderen Rahmen (die Dollar-Angaben sollten um die etwa 100-prozentige Inflation seit der Erhebung dieser Daten im Jahr 1984 bereinigt werden):

> Ein kleiner Copyshop hat einen Mitarbeiter, der dort seit sechs Monaten für 9 Dollar Stundenlohn arbeitet. Die Geschäfte laufen weiterhin zufriedenstellend, aber eine kleine Fabrik in der Nähe hat dichtgemacht, und die Arbeitslosigkeit ist gestiegen. Andere kleine Läden haben gerade zuverlässige Mitarbeiter für einen Stundenlohn von 7 Dollar eingestellt, um ähnliche Aufgaben zu erfüllen wie der Angestellte des Copyshops. Der Besitzer des Ladens kürzt den Lohn des Beschäftigten auf 7 Dollar.

Die befragten Personen billigten dies nicht: 83 Prozent beurteilten dieses Verhalten als »unfair« oder »sehr unfair«. Eine geringfügige Abwandlung der Frage verdeutlicht das Wesen der Verpflichtung des Arbeitgebers. Das Hintergrundszenario eines profitablen Ladens in einer Gegend mit hoher Arbeitslosigkeit ist das gleiche, aber jetzt

> kündigt der gegenwärtige Mitarbeiter, und der Besitzer beschließt, seinem Nachfolger 7 Dollar pro Stunde zu zahlen.

Eine große Mehrheit (73 Prozent) hielt diesen Schritt für »akzeptabel«. Offenkundig hat der Arbeitgeber keine moralische Pflicht, 9 Dollar pro Stunde zu zahlen. Der Anspruch ist personengebunden: Der gegenwärtige Mitarbeiter hat einen Anspruch darauf, seinen Lohn zu behalten, auch wenn es die Marktbedingungen dem Arbeitgeber erlauben würden, eine Lohnkürzung durchzusetzen. Der neue Mitarbeiter hat keinen Anspruch auf den Referenzlohn seines Vorgängers, und daher kann der Arbeitgeber den Lohn kürzen, ohne Gefahr zu laufen, als unfair gebrandmarkt zu werden.

Die Firma hat ihren eigenen legitimen Anspruch, der darin besteht, ihren gegenwärtigen Gewinn aufrechtzuerhalten. Wenn ihr ein Verlust droht, darf sie den Verlust auf andere abwälzen. Eine große Mehrheit der Befragten hält es nicht für unfair, wenn eine Firma bei sinkenden Gewinnen die Löhne ihrer Mitarbeiter kürzt. Wir beschrieben die Regeln so, dass sie zweifache Ansprüche definierten, nämlich sowohl die der Firma wie die der Personen, mit denen sie interagiert. Wenn sie selbst gefährdet ist, ist es von der Firma nicht unfair, sich egoistisch zu verhalten. Es wird von ihr nicht einmal erwartet, einen Teil der Verluste zu übernehmen; sie kann diese abwälzen.

Unterschiedliche Regeln legen fest, was ein Unternehmen tun darf, um seine Ertragslage zu verbessern oder einen Gewinnrückgang zu vermeiden. Wenn ein Unternehmen seine Produktionskosten senkt, gebieten es die Regeln der Fairness nicht, dass es den unerwarteten Geldsegen mit seinen Kunden oder seinen Mitarbeitern teilt. Selbstverständlich fanden die von uns befragten Personen eine Firma sympathischer und bezeichneten sie als fairer, wenn sie sich bei einem Gewinnanstieg großzügig zeigte, aber sie brandmarkten eine Firma, die nicht teilte, nicht als unfair. Sie zeigten sich nur dann empört, wenn eine Firma ihre Macht ausnutzte, um informelle

Verträge mit Mitarbeitern oder Kunden zu brechen und um anderen zum Zweck der Gewinnsteigerung einen Verlust aufzuerlegen. Die wichtige Aufgabe für Wissenschaftler, die ökonomische Fairness erforschen, besteht nicht darin, ideales Verhalten zu identifizieren, sondern die Linie zu finden, die akzeptables Verhalten von Handlungsweisen trennt, welche zu sozialer Brandmarkung und Bestrafung führen.

Wir waren nicht besonders optimistisch, als wir unseren Aufsatz über dieses Forschungsprojekt bei der *American Economic Review* zur Veröffentlichung einreichten. Unser Artikel stellte die damals herrschende ökonomische Lehrmeinung infrage, wonach wirtschaftliches Handeln von Eigennutz bestimmt wird und Fairnesserwägungen im Allgemeinen irrelevant sind. Wir stützten uns auch auf Erhebungsdaten, von denen Ökonomen im Allgemeinen wenig halten. Der Herausgeber des Journals sandte unseren Artikel zur Begutachtung jedoch zwei Ökonomen zu, die frei von solchen konventionellen Vorurteilen waren (wie erfuhren später ihre Identität; sie waren die wohlwollendsten, die der Herausgeber hätte finden können). Der Herausgeber wählte die richtigen Gutachter aus. Der Artikel wird häufig zitiert, und seine Schlussfolgerungen hatten langfristig Bestand. Neuere Forschungen haben die Beobachtungen zur referenzabhängigen Fairness bestätigt und darüber hinaus gezeigt, dass Fairnessaspekte ökonomisch bedeutsam sind – eine Tatsache, die wir vermutet, aber nicht bewiesen hatten.[14] Arbeitgeber, die gegen die Regeln der Fairness verstoßen, werden mit Produktivitätseinbußen bestraft, und Einzelhändler, die eine unfaire Preispolitik betreiben, müssen mit Umsatzeinbußen rechnen. Erfuhren Konsumenten aus einem neuen Katalog, dass der Einzelhändler jetzt für ein Produkt, das sie noch vor Kurzem zu einem höheren Preis gekauft hatten, weniger verlangte, kauften in Zukunft bei diesem Anbieter 15 Prozent weniger ein, was einem durchschnittlichen Verlust von 90 Dollar pro Kunde entspricht. Die Kunden haben offensichtlich den niedrigeren Preis als den Referenzpunkt wahrgenommen und glaubten von sich, einen Verlust erlitten zu haben, indem sie mehr bezahlt hatten, als angemessen war. Zudem waren die Kunden, die am stärksten reagierten, diejenigen, die mehr und teurere Artikel kauften. Die Verluste übertrafen bei Weitem die Gewinne aus den vermehrten Käufen, die durch die niedrigeren Preise im neuen Katalog hervorgerufen wurden.

In unfairer Weise Menschen Verluste aufzuerlegen kann riskant sein, wenn sich die Opfer revanchieren können. Zudem haben Experimente gezeigt, dass sich Fremde, die unfaires Verhalten beobachten, oftmals an der Bestrafung beteiligen. Neuroökonomen (Wissenschaftler, die die Ökonomik mit der Hirnforschung verknüpfen) haben die Gehirne von Menschen, die einen Fremden bestrafen, weil er sich unfair gegenüber einem anderen Fremden verhalten hat, mithilfe von Kernspintomografien untersucht. Bemerkenswerterweise geht altruistische Bestrafung mit einer gesteigerten Aktivität in den »Lustzentren« des Gehirns einher.[15] Es hat den Anschein, als wäre die Aufrechterhaltung der gesellschaftlichen Ordnung und der Fairnessregeln in dieser Weise ihre eigene Belohnung. Altruistische Bestrafung könnte durchaus der Kleber sein, der Gesellschaften zusammenhält. Doch unsere Gehirne sind nicht dafür ausgelegt, Großzügigkeit genauso zuverlässig zu belohnen, wie sie Geiz bestrafen. Auch hier finden wir wieder eine ausgeprägte Asymmetrie zwischen Verlusten und Gewinnen.

Der Einfluss von Verlustaversion und Ansprüchen erstreckt sich weit über den Bereich finanzieller Transaktionen hinaus. Juristen erkannten schon bald ihre Auswirkungen auf das Rechtswesen und die Rechtspflege. In einer Studie fanden David Cohen und Jack Knetsch viele Beispiele für einen deutlichen Unterschied zwischen tatsächlichen Verlusten und entgangenen Gewinnen bei gerichtlichen Entscheidungen.[16] So mag ein Einzelhändler, dessen Güter beim Transport verloren gegangen sind, für die Kosten, die ihm tatsächlich entstanden sind, entschädigt werden; für den entgangenen Gewinn dagegen wird er wahrscheinlich nicht entschädigt. Die geläufige Regel »Sei im Besitz, und du wohnst im Recht« bestätigt den moralischen Status des Referenzpunktes. In einer neueren Diskussion trägt Eyal Zamir das provokative Argument vor, dass die juristische Differenzierung zwischen Schadensersatz und Entschädigung für entgangenen Gewinn aufgrund ihrer asymmetrischen Effekte auf das individuelle Wohlergehen gerechtfertigt sei.[17] Wenn Menschen, die Vermögenseinbußen hinnehmen müssen, stärker geschädigt sind als Menschen, die lediglich keinen Gewinn erzielen, dann verdienen sie vielleicht auch einen stärkeren gesetzlichen Schutz.

## Zum Thema »Verluste«

»Diese Reform wird nicht durchkommen. Diejenigen, die Gefahr laufen, zu verlieren, werden härter kämpfen als diejenigen, die gewinnen könnten.«

»Jeder von ihnen glaubt, die Zugeständnisse des jeweils anderen wären weniger schmerzlich. Sie irren sich beide. Es ist nur die Asymmetrie der Verluste.«

»Es würde ihnen leichterfallen, das Abkommen neu auszuhandeln, wenn sie erkennen würden, dass der Kuchen größer wird. Sie verteilen keine Verluste; sie verteilen Gewinne.«

»Die Mieten in dieser Gegend sind in letzter Zeit gestiegen, aber unsere Mieter halten es nicht für fair, dass wir ihre Miete erhöhen. Sie glauben einen Anspruch auf ihre gegenwärtigen Konditionen zu haben.«

»Meine Kunden nehmen mir die Preiserhöhung nicht übel, weil sie wissen, dass auch meine Kosten gestiegen sind. Sie erkennen mein Recht an, weiterhin mit Gewinn zu arbeiten.«

# *29. Das viergeteilte Muster*

Immer wenn Sie eine globale Beurteilung eines komplexen Objektes vornehmen – eines Autos, das Sie vielleicht kaufen werden, Ihres Schwiegersohns oder einer ungewissen Situation –, gewichten Sie dessen Merkmale. Dies ist schlicht eine umständliche Art, zu sagen, dass manche Merkmale unser Urteil stärker beeinflussen als andere. Die Gewichtung geschieht unabhängig davon, ob wir uns dessen bewusst sind; sie ist eine Operation von System 1. Ihre Gesamtbeurteilung eines Autos kann den niedrigen Kraftstoffverbrauch, den Komfort oder das Design unterschiedlich stark gewichten. Ihr Urteil über Ihren Schwiegersohn mag mehr oder weniger stark davon abhängig sein, wie reich, gut aussehend oder zuverlässig er ist. In ähnlicher Weise schreiben Sie bei der Einschätzung unsicherer zukünftiger Ereignisse möglichen Szenarien unterschiedliche Gewichte bei. Die Gewichtungen korrelieren zweifellos mit den Wahrscheinlichkeiten der Ergebnisse: Eine 50-prozentige Wahrscheinlichkeit, eine Million zu gewinnen, ist viel attraktiver als eine 1-prozentige Wahrscheinlichkeit, die gleiche Summe zu gewinnen. Die Gewichtung erfolgt manchmal bewusst und wohlüberlegt. Sehr oft aber sind wir bloße Beobachter einer globalen Bewertung, die unser System 1 vornimmt.

## Wahrscheinlichkeiten verändern

Ein Grund für die Beliebtheit der Glücksspielmetapher bei der Erforschung der Entscheidungsfindung besteht darin, dass sie eine natürliche Regel für die Gewichtung der Ergebnisse einer Gewinnaussicht *(prospect)* bereitstellt: Je wahrscheinlicher ein Ergebnis ist, umso stärker sollte es gewichtet werden. So beträgt zum Beispiel der erwartete Nutzen einer »20-prozentigen Chance, 1000 Dollar zu gewinnen, und einer 75-prozentigen Chance, 100 Dollar zu gewinnen« 275 Dollar. In der Zeit vor Bernoulli wurden Lotterien nach ihrem Erwartungswert beurteilt. Bernoulli behielt diese Methode der Gewichtung von Ergebnissen bei – Erwartungsprinzip genannt –, wandte sie jedoch auf den psychologischen Wert der Ergebnisse an. Der Nutzen einer Lotterie ist in dieser Theorie der Mittelwert der Nutzwerte ihrer – mit ihren jeweiligen Wahrscheinlichkeiten gewichteten – Ergebnisse.

Das Erwartungsprinzip liefert keine zutreffende Beschreibung dessen, wie wir die Wahrscheinlichkeiten, die mit risikobehafteten Gewinnaussichten verbunden sind, einschätzen. In den vier nachfolgenden Beispielen erhöhen sich unsere Chancen, 1 Million Dollar zu erhalten, um 5 Prozent. Ist diese Neuigkeit in jedem Fall gleich nützlich?

a. Von 0 auf 5 Prozent
b. Von 5 auf 10 Prozent
c. Von 60 auf 65 Prozent
d. Von 95 auf 100 Prozent

Das Erwartungsprinzip besagt, dass sich unser Nutzen in jedem Fall um genau 5 Prozent des Nutzens erhöht, der mit 1 Million Dollar verbunden ist. Beschreibt diese Vorhersage unsere Erfahrungen? Natürlich nicht. Jeder würde zustimmen, dass 0 → 5 Prozent und 95 → 100 Prozent eindrucksvoller sind als entweder 5 → 10 Prozent oder 60 → 65 Prozent. Die Erhöhung der Wahrscheinlichkeit von 0 auf 5 Prozent verändert die Situation grundlegend, schafft eine Möglichkeit, die zuvor nicht existierte, eine Hoffnung, den Preis zu gewinnen. Es ist eine qualitative Veränderung, während 5 → 10 Prozent nur eine quantitative Verbesserung ist. Die Veränderung von 5 auf 10 Prozent verdoppelt die

Gewinnwahrscheinlichkeit, aber es besteht allgemeines Einvernehmen darüber, dass sich der psychologische Wert der Gewinnerwartung nicht verdoppelt. Die starke Wirkung von 0 → 5 Prozent verdeutlicht den Möglichkeitseffekt, der dafür sorgt, dass höchst unwahrscheinliche Ergebnisse unverhältnismäßig stärker gewichtet werden, als sie es »verdienen«. Menschen, die Lotterielose in sehr großer Zahl kaufen, zeigen sich bereit, für sehr kleine Chancen, einen großen Preis zu gewinnen, viel mehr als den erwarteten Nutzen zu zahlen.

Die Verbesserung von 95 auf 100 Prozent ist eine weitere qualitative Veränderung mit weitreichenden Folgen, der Sicherheitseffekt. Ergebnisse, die fast sicher sind, werden geringer gewichtet, als es ihre Wahrscheinlichkeit rechtfertigt. Um die Bedeutung des Sicherheitseffekts zu verstehen, stellen Sie sich vor, dass Sie 1 Million Dollar geerbt haben, aber Ihre habgierige Stiefschwester hat das Testament vor Gericht angefochten. Die Entscheidung wird morgen erwartet. Ihr Anwalt versichert Ihnen, Sie hätten das Recht auf Ihrer Seite und würden mit einer Wahrscheinlichkeit von 95 Prozent gewinnen, aber er erinnert Sie auch daran, dass Gerichtsurteile nie hundertprozentig vorhersagbar sind. Jetzt tritt eine Risikoübernahme-Gesellschaft an Sie heran und bietet an, Ihnen den Fall gegen sofortige Zahlung von 910 000 Dollar abzukaufen – Sie müssen sich sofort entscheiden. Das Angebot ist niedriger (um 40 000 Dollar) als der Erwartungswert des Urteils (der 950 000 Dollar beträgt), aber sind Sie sicher, das Angebot ablehnen zu wollen? Wenn sich ein solches Ereignis tatsächlich in Ihrem Leben zutragen sollte, sollten Sie wissen, dass es eine eigene Branche für »strukturierte Abwicklungen« gibt, die zu einem saftigen Preis Sicherheit bieten, indem sie sich den Sicherheitseffekt zunutze machen.

Möglichkeit und Sicherheit haben auf dem Gebiet von Verlusten ähnlich starke Effekte. Wenn eine nahestehende Person operiert wird, ist ein 5-prozentiges Risiko einer Amputation sehr schlecht – viel mehr als halb so schlecht wie ein 10-prozentiges Risiko. Wegen des Möglichkeitseffekts neigen wir dazu, kleine Risiken überzubewerten, und wir sind bereit, viel mehr als den Erwartungswert zu bezahlen, um sie auszuschalten. Der psychologische Unterschied zwischen einem 95-prozentigen Katastrophenrisiko und der Gewissheit, dass es zu einer Katastrophe kommen wird, scheint sogar noch größer zu sein; der Hoffnungsschimmer, dass die Katastrophe ausbleibt, hat eine sehr große Bedeutung. Die

Übergewichtung geringer Wahrscheinlichkeiten steigert die Attraktivität von Glücksspielen und Versicherungspolicen.

Die Schlussfolgerung ist einfach: Die Entscheidungsgewichte, die Menschen Ergebnissen zuschreiben, sind, im Widerspruch zum Erwartungsprinzip, nicht identisch mit den Wahrscheinlichkeiten dieser Ergebnisse. Unwahrscheinliche Ergebnisse werden übergewichtet – dies ist der Möglichkeitseffekt. Ergebnisse, die fast sicher sind, werden im Verhältnis zu ihrer tatsächlichen Eintrittssicherheit untergewichtet. Das Erwartungsprinzip, wonach Nutzwerte nach ihrer Wahrscheinlichkeit gewichtet werden, ist schlechte Psychologie.

Aber nun wird es ziemlich kompliziert, weil es ein starkes Argument dafür gibt, dass sich ein Entscheider, der rational sein will, dem Erwartungsprinzip fügen *muss*. Dies war der Hauptpunkt der axiomatischen Version der Nutzentheorie, die von Neumann und Morgenstern im Jahr 1944 vorgestellt wurde. Sie bewiesen, dass jede Gewichtung unsicherer Ergebnisse, die nicht streng proportional zu ihrer Wahrscheinlichkeit ist, zu Inkonsistenzen und anderen Katastrophen führt.[1] Ihre Ableitung des Erwartungsprinzips aus Axiomen der rationalen Wahl wurde sofort als eine monumentale Leistung anerkannt, die die Erwartungsnutzentheorie ins Zentrum des Modells vom rationalen Agenten in der Ökonomik und anderen Sozialwissenschaften stellte. Als mich Amos dreißig Jahre später mit ihrer Arbeit vertraut machte, stellte er sie mir als einen Gegenstand ehrfürchtiger Verehrung vor. Er machte mich auch mit einer berühmten Herausforderung dieser Theorie bekannt.

## Das Allais-Paradoxon

Im Jahr 1952, einige Jahre nach der Veröffentlichung der Theorie von Neumann und Morgenstern, wurde eine Konferenz in Paris einberufen, auf der ökonomische Aspekte des Risikobegriffs diskutiert werden sollten. Viele der berühmtesten Ökonomen der damaligen Zeit nahmen daran teil. Unter den amerikanischen Gästen waren die zukünftigen Nobelpreisträger Paul Samuelson, Kenneth Arrow und Milton Friedman sowie der bedeutende Statistiker Jimmie Savage.

Einer der Organisatoren der Pariser Konferenz war Maurice Allais, der einige Jahre später ebenfalls mit dem Nobelpreis ausgezeichnet werden sollte. Allais hatte etwas in petto, eine Reihe von Fragen über Wahlhandlungen, vor die er seine berühmten Zuhörern stellte. In der Terminologie dieses Kapitels formuliert, wollte Allais zeigen, dass seine Gäste anfällig seien für einen Sicherheitseffekt und daher gegen die Erwartungsnutzentheorie und die Axiome rationaler Entscheidung, auf denen diese Theorie gründet, verstießen. Die folgende Reihe von Wahlmöglichkeiten ist eine vereinfachte Version des von Allais konstruierten Problems.[2] Für welche der beiden Optionen bei den Problemen A und B würden Sie sich entscheiden?

a. Eine 61-prozentige Chance, 520 000 Dollar zu gewinnen, ODER eine 63-prozentige Chance, 500 000 Dollar zu gewinnen.
b. Eine 98-prozentige Chance, 520 000 Dollar zu gewinnen, ODER eine 100-prozentige Chance, 500 000 Dollar zu gewinnen.

Wenn Sie wie die meisten anderen Menschen sind, würden Sie die obere Option bei Problem A und die untere Option bei Problem B vorziehen. Wenn dies Ihre Präferenzen waren, dann haben Sie gerade eine logische Sünde begangen und gegen die Regeln rationaler Entscheidung verstoßen. Die in Paris versammelten illustren Ökonomen begingen ähnliche Sünden in einer komplizierteren Version des Allais-Paradoxons.

Um zu verstehen, weshalb diese Wahlen problematisch sind, stellen Sie sich vor, dass das Ergebnis durch »blindes« Ziehen aus einer Urne mit hundert Murmeln bestimmt wird – Sie gewinnen, wenn Sie eine rote Murmel ziehen, Sie verlieren, wenn Sie eine weiße ziehen. Bei Problem A

präferieren fast alle die erste Urne, obwohl sie weniger siegreiche rote Murmeln enthält, weil die Differenz in der Höhe des Gewinns eindrucksvoller ist als die Differenz bei den Gewinnchancen. Bei Problem B entscheidet sich eine große Mehrheit für die Urne, die einen Gewinn von 500 000 Dollar garantiert. Außerdem fühlen sich Menschen mit beiden Optionen wohl – bis ihnen die Logik des Problems verdeutlicht wird.

Vergleichen Sie die beiden Probleme, und Sie werden sehen, dass die beiden Urnen von Problem B günstigere Versionen der Urnen von Problem A sind, wobei in jeder Urne 37 weiße Murmeln durch gewinnbringende rote Murmeln ersetzt wurden. Die Verbesserung in der ersten Urne ist eindeutig größer als die Verbesserung in der zweiten, da jede rote Murmel Ihnen eine Chance gibt, in der ersten Urne 520 000 Dollar zu gewinnen, während Sie in der zweiten Urne nur 500 000 Dollar gewinnen können. Also begannen Sie beim ersten Problem mit einer Präferenz für die erste Urne, die anschließend stärker verbessert wurde als die zweite Urne – aber jetzt ziehen Sie die zweite Urne vor! Dieses Entscheidungsmuster ist logisch nicht sinnvoll, aber es lässt sich leicht psychologisch erklären: hier greift der Sicherheitseffekt. Der 2-prozentige Unterschied zwischen einer 100-prozentigen und einer 98-prozentigen Gewinnchance bei Problem B ist viel eindrucksvoller als derselbe Unterschied zwischen 63 und 61 Prozent bei Problem A.

Wie von Allais vorhergesehen, bemerkten die hochkarätigen Teilnehmer der Konferenz nicht, dass ihre Präferenzen gegen die Nutzentheorie verstießen, bis er sie am Ende der Konferenz auf die Tatsache aufmerksam machte. Allais wollte, dass diese Ankündigung wie eine Bombe einschlägt: Die führenden Entscheidungstheoretiker der Welt hatten Präferenzen, die im Widerspruch zu ihrem eigenen Rationalitätskonzept standen! Er glaubte offenbar, dies würde seine Zuhörer veranlassen, den theoretischen Ansatz aufzugeben, den er recht verächtlich »die amerikanische Denkschule« nannte, und eine alternative Logik der Entscheidung zu übernehmen, die er entwickelt hatte. Er sollte schwer enttäuscht werden.[3]

Ökonomen, die keine Fans der Entscheidungstheorie waren, ignorierten zum größten Teil das Allais-Problem. Wie es häufig geschieht, wenn eine Theorie, die allgemein anerkannt ist und sich als nützlich erwiesen hat, in Zweifel gezogen wird, taten sie das Problem als eine Anomalie ab und wandten weiterhin die Erwartungsnutzentheorie an, als wäre nichts geschehen. Dagegen haben Entscheidungstheoretiker – eine gemischte

Gruppe von Statistikern, Wirtschaftswissenschaftlern, Philosophen und Psychologen – Allais' Herausforderung sehr ernst genommen. Als Amos und ich mit unserer Zusammenarbeit begannen, war eines unserer ersten Ziele, eine befriedigende psychologische Erklärung des Allais-Paradoxons zu finden.

Die meisten Entscheidungstheoretiker, bemerkenswerterweise Allais eingeschlossen, glaubten weiterhin an die Rationalität des Menschen und versuchten, die Regeln rationaler Entscheidungsfindung so zu modifizieren, dass das Allais-Muster zulässig wird. Im Lauf der Jahre gab es mehrere Versuche, eine plausible Begründung für den Sicherheitseffekt zu finden, keiner davon war besonders überzeugend. Amos hatte für diese Bemühungen wenig Verständnis; er nannte die Theoretiker, die Verstöße gegen die Nutzentheorie rational erklären wollten, »Anwälte für die Fehlgeleiteten«. Wir schlugen eine andere Richtung ein und behielten die Nutzentheorie als eine Logik rationaler Wahl zwar bei, gaben jedoch die Annahme auf, Menschen wären vollkommen rationale Entscheider. Wir setzten es uns als Aufgabe, eine psychologische Theorie zu entwickeln, die die Entscheidungen von Menschen unabhängig davon, ob sie rational sind, beschreiben sollte. In der Neuen Erwartungstheorie sollten die Entscheidungsgewichte nicht identisch sein mit den Wahrscheinlichkeiten.

## Entscheidungsgewichte

Viele Jahre nach der Veröffentlichung der Neuen Erwartungstheorie führten Amos und ich eine Studie durch, in der wir die Entscheidungsgewichte, die individuelle Präferenzen für Glücksspiele erklären, mithilfe niedriger Geldeinsätze maßen. Die geschätzten Gewinne sind in Tabelle 4 aufgeführt.[4]

| Wahrscheinlichkeit (Prozent) | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 0 | 1 | 2 | 5 | 10 | 20 | 50 | 80 | 90 | 95 | 98 | 99 | 100 |
| Entscheidungsgewicht | | | | | | | | | | | | |
| 0 | 5.5 | 8.1 | 13.2 | 18.6 | 26.1 | 42.1 | 60.1 | 71.2 | 79.3 | 87.1 | 91.2 | 100 |

**Tabelle 4**

Sie können sehen, dass die Entscheidungsgewichte identisch sind mit den entsprechenden Wahrscheinlichkeiten an den Extremen: Beide sind gleich null, wenn das Ergebnis unmöglich ist, und beide sind gleich hundert, wenn das Ergebnis sicher ist. Aber in der Nähe dieser Punkte weichen die Entscheidungsgewichte deutlich von den Wahrscheinlichkeiten ab. Am unteren Ende begegnen wir dem Möglichkeitseffekt; unwahrscheinliche Ereignisse sind stark übergewichtet. So beträgt zum Beispiel das Entscheidungsgewicht, das einer Wahrscheinlichkeit von 2 Prozent entspricht, 8,1. Wenn Menschen den Axiomen des rationalen Entscheidungsverhaltens gehorchten, wäre das Entscheidungsgewicht zwei – also wird das seltene Ereignis um einen Faktor von vier übergewichtet. Der Sicherheitseffekt am anderen Ende der Wahrscheinlichkeitsskala ist noch frappierender. Ein 2-prozentiges Risiko, die Prämie *nicht* zu gewinnen, verringert den Nutzen des Glücksspiels um 13 Prozent, von 100 auf 87,1.

Um die Asymmetrie zwischen dem Möglichkeitseffekt und dem Sicherheitseffekt zu verstehen, stellen Sie sich zunächst einmal vor, dass Sie eine Chance von 1 Prozent haben, 1 Million Dollar zu gewinnen. Sie werden das Ergebnis morgen erfahren. Stellen Sie sich jetzt vor, dass Sie fast sicher sind, 1 Million Dollar zu gewinnen, aber dass Sie mit einer

Wahrscheinlichkeit von 1 Prozent leer ausgehen. Auch hier erfahren Sie das Ergebnis morgen. In der zweiten Situation scheint die Angst markanter zu sein als die Hoffnung in der ersten. Der Sicherheitseffekt ist auch deutlicher ausgeprägt als der Möglichkeitseffekt, wenn das Ergebnis ein chirurgisches Desaster statt eines finanziellen Gewinns ist. Vergleichen Sie die Intensität, mit der Sie sich auf den schwachen Hoffnungsschimmer bei einer höchstwahrscheinlich tödlich verlaufenden Operation konzentrieren, verglichen mit der Furcht vor einem 1-prozentigen Risiko.

Die Kombination des Sicherheitseffekts und des Möglichkeitseffekts an den beiden Enden der Wahrscheinlichkeitsskala geht zwangsläufig mit einer unzulänglichen Sensitivität für dazwischenliegende Wahrscheinlichkeiten einher. Sie können sehen, dass das Spektrum der Wahrscheinlichkeiten zwischen 5 und 95 Prozent mit einem viel schmaleren Bereich von Entscheidungsgewichten (von 13,2 bis 79,3) verbunden ist, was etwa zwei Dritteln dessen entspricht, was rational zu erwarten ist. Neurowissenschaftler haben diese Beobachtungen bestätigt und Hirnareale entdeckt, die auf Veränderungen der Wahrscheinlichkeit, einen Preis zu gewinnen, reagieren. Die Reaktion des Gehirns auf Wahrscheinlichkeitsschwankungen weist bemerkenswerte Übereinstimmungen mit den Entscheidungsgewichten auf, die auf der Basis empirischer Wahlen geschätzt wurden.[5]

Extrem niedrige oder hohe Wahrscheinlichkeiten (unter 1 Prozent oder über 99 Prozent) sind ein Sonderfall. Es ist schwierig, sehr seltenen Ereignissen ein bestimmtes Entscheidungsgewicht zuzuschreiben, weil sie manchmal gänzlich ignoriert werden und ihnen faktisch ein Entscheidungsgewicht von null zugeschrieben wird. Andererseits wird man diese sehr seltenen Ereignisse, wenn man sie nicht ignoriert, mit Sicherheit übergewichten. Die meisten von uns verbringen sehr wenig Zeit damit, sich den Kopf über eine mögliche Kernschmelze in einem Reaktor zu zerbrechen oder über große Erbschaften von unbekannten Verwandten zu fantasieren. Doch wenn wir unsere Aufmerksamkeit auf ein unwahrscheinliches Ereignis konzentrieren, gewichten wir es viel stärker, als es seine Wahrscheinlichkeit verdient. Zudem haben wir wenig Gespür für Risikoschwankungen bei niedrigen Wahrscheinlichkeiten. Ein Krebsrisiko von 0,001 Prozent lässt sich nicht leicht von einem Risiko von 0,00001 Prozent unterscheiden, obgleich Ersteres 3000

Krebsneuerkrankungen im Jahr entsprechen würde, während Letzteres nur dreißig neue Krebsfälle bedeutete.

Wenn man einer Bedrohung Aufmerksamkeit schenkt, ist man besorgt – und in den Entscheidungsgewichten spiegelt sich das Ausmaß der Besorgtheit wider. Wegen des Möglichkeitseffekts ist die Sorge nicht proportional zur Wahrscheinlichkeit der Bedrohung. Es genügt nicht, das Risiko zu verringern oder abzuschwächen; um die Sorge zu beseitigen, muss die Wahrscheinlichkeit auf null reduziert werden.

Die nachstehende Frage ist aus einer Studie über die Rationalität der Bewertungen von Gesundheitsrisiken durch Verbraucher übernommen, die in den 1980er-Jahren von einem Team von Ökonomen veröffentlicht wurde. Die Umfrage richtete sich an Eltern kleiner Kinder.[6]

> Angenommen, Sie verwenden gegenwärtig ein Insektenspray, das Sie pro Dose 10 Dollar kostet und das zu 15 Inhalationsvergiftungen und 15 Vergiftungen bei Kindern je 10 000 benutzte Dosen Insektenspray führt.
>
> Sie hören, dass es ein teureres Insektizid gibt, das beide Risiken auf 5 für jeweils 10 000 Dosen reduziert. Wie viel wären Sie bereit dafür zu zahlen?

Die Eltern waren bereit, im Schnitt zusätzliche 2,38 Dollar zu zahlen, um das Risiko um zwei Drittel zu verringern, von 15 auf 5 je 10 000 Dosen. Um es vollständig auszuschalten, waren sie bereit, 8,09 Dollar zu bezahlen, also mehr als dreimal so viel. Weitere Fragen zeigten, dass die Eltern die beiden Risiken (Inhalation und Kindervergiftung) als getrennte Sorgen behandelten und bereit waren, eine Sicherheitsprämie für die vollständige Beseitigung beider zu zahlen. Diese Prämie ist kompatibel mit der Psychologie der Sorge, aber nicht mit dem rationalen Modell.[7]

## Das viergeteilte Muster

Als Amos und ich unsere Arbeit an der Neuen Erwartungstheorie begannen, gelangten wir schnell zu zwei Schlussfolgerungen: Menschen messen weniger dem Vermögen an sich als vielmehr Gewinnen und Verlusten einen Wert bei, und die Entscheidungsgewichte, die sie Ergebnissen zuschreiben, unterscheiden sich von Wahrscheinlichkeiten. Keine der beiden Annahmen war völlig neu, aber beide zusammen erklärten ein spezifisches Muster von Präferenzen, das wir das »viergeteilte Muster« nannten. Die Bezeichnung blieb erhalten. Die Szenarien sind unten verdeutlicht:

| | GEWINNE | VERLUSTE |
|---|---|---|
| HOHE WAHRSCHEIN-LICHKEIT | 95-prozentige Wahrscheinlichkeit, 10 000 Dollar zu gewinnen | 95-prozentige Wahrscheinlichkeit, 10 000 Dollar zu verlieren |
| Sicherheitseffekt | Angst vor Enttäuschung | Hoffnung, Verluste zu vermeiden |
| | RISIKOSCHEU | RISIKOFREUDIG |
| | Annahme eines ungünstigen Vergleichs | Ablehnung eines günstigen Vergleichs |
| GERINGE WAHRSCHEIN-LICHKEIT | 5-prozentige Wahrscheinlichkeit, 10 000 Dollar zu gewinnen | 5-prozentige Wahrscheinlichkeit, 10 000 Dollar zu verlieren |
| Möglichkeits-effekt | Hoffnung auf einen hohen Gewinn | Furcht vor einem hohen Verlust |
| | RISIKOFREUDIG | RISIKOSCHEU |
| | Ablehnung eines günstigen Vergleichs | Annahme eines ungünstigen Vergleichs |

**Abbildung 13**

– Die oberste Reihe in jeder Zelle zeigt eine illustrative Gewinn-/Verlustaussicht.
– Die zweite Reihe charakterisiert die zentrale Emotion, die die Aussicht *(prospect)* hervorruft.

– Die dritte Reihe gibt an, wie sich die meisten Menschen verhalten, wenn sie die Wahl zwischen einer Lotterie und einem sicheren Gewinn (oder Verlust) haben, der ihrem Erwartungswert entspricht (zum Beispiel zwischen »einer 95-prozentigen Chance, 10 000 Dollar zu gewinnen« und »9500 Dollar sicher erhalten«). Entscheidungen gelten als risikoscheu, wenn die sichere Option vorgezogen wird, und als risikofreudig, wenn die Lotterie bevorzugt wird.
– Die vierte Reihe beschreibt die erwarteten Einstellungen eines Beklagten und eines Klägers, wenn sie über einen Vergleich in einem Zivilprozess verhandeln.

Das viergeteilte Muster der Präferenzen gilt als eine der zentralen Errungenschaften der Neuen Erwartungstheorie. Drei der vier Zellen waren bekannt; die vierte (oben rechts) war neu und unerwartet.

– Die obere linke Zelle war diejenige, die von Bernoulli diskutiert wurde: Menschen sind risikoscheu, wenn sie Aussichten mit einer hohen Wahrscheinlichkeit eines großen Gewinns erwägen. Sie sind bereit, sich mit weniger als dem Erwartungswert eines Glücksspiels zufriedenzugeben, um einen sicheren Gewinn festzuschreiben.
– Der Möglichkeitseffekt im unteren linken Feld erklärt, weshalb sich Lotterien so großer Beliebtheit erfreuen. Wenn der Hauptgewinn sehr groß ist, scheinen Loskäufer gleichgültig zu sein gegenüber der Tatsache, dass ihre Gewinnchance verschwindend gering ist. Ein Lotterielos ist das ultimative Beispiel des Möglichkeitseffekts. Ohne ein Los kann man nicht gewinnen, mit einem Los hat man eine Chance, und ob die Chance äußerst gering oder nur gering ist, spielt keine große Rolle. Selbstverständlich erwerben Menschen mit einem Los mehr als eine Gewinnchance; sie erwerben das Recht, sich Wunschträumen über einen Gewinn hinzugeben.
– Die Zelle unten rechts beschreibt die Situation, in der man sich gegen ein Risiko versichert. Menschen sind bereit, für eine Versicherung weit mehr zu zahlen, als dem Erwartungswert des Risikos entspricht – auf diese Weise decken Versicherungsgesellschaften ihre Kosten und machen ihre Gewinne. Auch hier kaufen sich Menschen mehr als den Schutz gegen eine unwahrscheinliche Katastrophe; sie befreien sich von einer Sorge und erwerben sich eine Seelenruhe.

Die Ergebnisse für das Feld oben rechts haben uns zunächst überrascht. Wir waren es gewohnt, in Kategorien der Risikoaversion zu denken, einmal abgesehen von der Zelle links unten, wo Lotterien bevorzugt werden. Als wir unsere Entscheidungen für schlechte Optionen analysierten, erkannten wir schnell, dass wir im Feld der Verluste genauso risikofreudig waren wie im Feld der Gewinne risikoscheu. Wir waren nicht die Ersten, die Risikofreude bei negativen Gewinnaussichten beobachteten – mindestens zwei Autoren hatten darüber berichtet, aber keine weiteren Konsequenzen daraus gezogen.[8] Aber wir hatten das Glück, jetzt über einen Bezugsrahmen zu verfügen, der den Befund der Risikofreude leicht interpretierbar machte, und dies war ein Meilenstein in unserem Denken. Tatsächlich identifizierten wir zwei Ursachen für diesen Effekt.

Erstens nimmt die Empfindlichkeit ab. Der sichere Verlust ist sehr aversiv, weil die Reaktion auf einen Verlust von 900 Dollar mehr als 90 Prozent so intensiv ist wie die Reaktion auf einen Verlust von 1000 Dollar. Der zweite Faktor ist vielleicht noch stärker: Das Entscheidungsgewicht, das einer Wahrscheinlichkeit von 90 Prozent entspricht, beträgt nur etwa 71, ist also viel niedriger als die Wahrscheinlichkeit. Das Ergebnis ist, dass man bei einer Wahl zwischen einem sicheren Verlust und einer Lotterie mit einer hohen Wahrscheinlichkeit eines größeren Verlusts aufgrund der abnehmenden Empfindlichkeit eine stärkere Aversion gegen den sicheren Verlust hat, während der Sicherheitseffekt die Aversion gegenüber der Lotterie reduziert. Dieselben beiden Faktoren steigern die Attraktivität der sicheren Option und verringern die Attraktivität des Glücksspiels, wenn die Ergebnisse positiv sind.

Der Verlauf der Wertfunktion und die Entscheidungsgewichte tragen beide zu dem Muster bei, das in der oberen Reihe von Abbildung 13 zu beobachten ist. In der unteren Reihe dagegen wirken die beiden Faktoren in entgegengesetzte Richtungen; die abnehmende Empfindlichkeit begünstigt weiterhin eine Risikoscheu für Gewinne und eine Risikofreude für Verluste, aber die Übergewichtung niedriger Wahrscheinlichkeiten überwindet diesen Effekt und erzeugt das beobachtete Muster eines riskanten Spielens um Gewinne und einer Vermeidung von Verlusten.

In der oberen rechten Zelle ereignen sich viele unerfreuliche Situationen. Hier entscheiden sich Menschen, die mit sehr negativen Optionen konfrontiert sind, für wenig aussichtsreiche Wetten, wobei sie für eine geringe Hoffnung, einen großen Verlust zu vermeiden, eine hohe

Wahrscheinlichkeit in Kauf nehmen, alles noch schlimmer zu machen. Eine solche Risikobereitschaft macht aus beherrschbaren Misserfolgen Katastrophen. Der Gedanke, den hohen sicheren Verlust einfach hinzunehmen, ist zu schmerzlich und die Hoffnung auf eine vollständige Entlastung so verlockend, dass man nicht die vernünftige Entscheidung trifft, die eigenen Verluste zu begrenzen. Hier verschleudern Unternehmen, die aufgrund der überlegenen Technologie eines Wettbewerbers Boden verlieren, ihre verbliebenen Vermögenswerte in dem vergeblichen Bemühen, zu dem Wettbewerber aufzuschließen. Weil Niederlagen so schwer zu verkraften sind, kämpfen die Verlierer in Kriegen oftmals weit über den Punkt hinaus, an dem der Sieg der anderen Seite sicher und nur eine Frage der Zeit ist.

## Glücksspiele im Schatten des Gesetzes

Der Rechtswissenschaftler Chris Guthrie hat das viergeteilte Muster auf zwei Situationen angewandt, in denen der Kläger und der Beklagte in einem Zivilprozess einen möglichen Vergleich in Erwägung ziehen – mit eindrucksvollen Ergebnissen. Die Situationen unterscheiden sich bezüglich der Erfolgsaussichten der Klage.

Wie in einem Szenario, das wir weiter vorn betrachteten, sind Sie ein Kläger in einem Zivilprozess, in dem Sie eine hohe Summe Schadensersatz fordern. Der Prozess verläuft sehr gut, und Ihr Anwalt zitiert die Meinung von Experten, wonach Sie eine 95-prozentige Chance haben, einen uneingeschränkten Sieg davonzutragen. Aber er fügt diese Warnung hinzu: »Man kennt das Ergebnis immer erst, wenn der Richter das Urteil verliest.« Ihr Anwalt drängt Sie, einen Vergleich anzunehmen, bei dem Sie vielleicht nur 90 Prozent der von Ihnen gerichtlich geforderten Schadensersatzsumme erhalten. Sie befinden sich im oberen linken Feld des viergeteilten Musters, und die Frage, die Ihnen durch den Kopf geht, lautet: »Bin ich bereit, eine wenn auch geringe Wahrscheinlichkeit, völlig leer auszugehen, in Kauf zu nehmen? Selbst 90 Prozent der geforderten Summe sind eine Menge Geld, und ich bekomme sie sofort.« Zwei Emotionen werden ausgelöst, die beide in die gleiche Richtung drängen: die Anziehungskraft eines sicheren (und beträchtlichen) Gewinns und die Furcht vor einer starken Enttäuschung und großem Bedauern, wenn Sie einen Vergleich ablehnen und den Prozess verlieren. Sie spüren einen starken Druck, der in einer solchen Situation in der Regel zu umsichtigem Verhalten führt. Ein Kläger mit einer gut fundierten Klage wird vermutlich risikoscheu sein.

Versetzen Sie sich jetzt im selben Prozess in den Beklagten. Obwohl Sie die Hoffnung auf eine Entscheidung zu Ihren Gunsten noch nicht ganz aufgegeben haben, erkennen Sie, dass der Prozess schlecht für sie läuft. Die Anwälte des Klägers haben einen Vergleich vorgeschlagen, bei dem Sie 90 Prozent der ursprünglich geforderten Summe zahlen müssten, und es ist klar, dass sie sich nicht mit weniger zufriedengeben werden. Werden Sie dem Vergleich zustimmen oder den Prozess fortsetzen? Weil Sie mit hoher Wahrscheinlichkeit einen Verlust hinnehmen müssen, gehört Ihre Situation in die obere rechte Zelle. Die Versuchung, weiterzukämpfen, ist groß; der

Vergleich, den der Kläger angeboten hat, ist fast genauso schmerzlich wie das denkbar schlimmste Ergebnis, und es besteht noch immer Hoffnung, dass Sie vor Gericht obsiegen werden. Auch hier sind zwei Emotionen im Spiel: Der sichere Verlust ist abstoßend, und die Möglichkeit, den Prozess zu gewinnen, ist höchst attraktiv. Ein Beklagter mit geringen Erfolgsaussichten vor Gericht wird vermutlich risikofreudig sein, also bereit sein, zu pokern, statt einen sehr ungünstigen Vergleich zu akzeptieren. In der Konfrontation zwischen einem risikoscheuen Kläger und einem risikofreudigen Beklagten hat der Beklagte die besseren Karten. Die überlegene Verhandlungsposition des Beklagten sollte sich in dem ausgehandelten Vergleich widerspiegeln, wobei sich der Kläger mit weniger als der statistisch zu erwartenden Entschädigungssumme zufriedengibt. Diese Vorhersage aus dem viergeteilten Muster wurde bei Experimenten mit Jurastudenten und praktizierenden Richtern und auch durch Analysen tatsächlicher Verhandlungen im Rahmen von Zivilprozessen bestätigt.[9]

Betrachten wir jetzt den Fall einer »mutwilligen Klage«, wo ein Kläger auf dürftiger Anspruchsgrundlage eine hohe Schadensersatzforderung stellt, die wahrscheinlich vor Gericht scheitern wird. Beide Seiten kennen die Wahrscheinlichkeiten, und beide wissen, dass der Kläger bei einem Vergleich nur einen kleinen Teil des gerichtlich geforderten Schadensersatzes bekommen wird. Die Verhandlung wird in der unteren Zeile des viergeteilten Musters geführt. Der Kläger ist in der linken Zelle, mit einer geringen Wahrscheinlichkeit, eine sehr große Summe zu gewinnen; die mutwillige Schadensersatzforderung ist ein Lotterielos für eine sehr große Gewinnsumme.[10] Die Übergewichtung der niedrigen Erfolgswahrscheinlichkeit ist in dieser Situation ganz natürlich und veranlasst den Kläger dazu, in den Verhandlungen forsch und aggressiv aufzutreten. Für den Beklagten ist der Prozess ein Ärgernis mit einem geringen Risiko eines sehr schlechten Ergebnisses. Die Übergewichtung der geringen Wahrscheinlichkeit eines hohen Verlusts begünstigt die Risikoscheu, das Schließen eines Vergleichs über eine bescheidene Summe entspricht dem Kauf einer Versicherung gegen das unwahrscheinliche Ereignis eines negativen Gerichtsurteils. Der Spieß hat sich jetzt umgedreht: Der Kläger ist bereit, etwas zu riskieren, während der Beklagte auf Nummer sicher gehen will. Kläger mit schlecht fundierten Ansprüchen

erzielen wahrscheinlich einen großzügigeren Vergleich, als es die statistischen Merkmale der Situation rechtfertigen würden.

Die in dem viergeteilten Muster beschriebenen Entscheidungen sind nicht offensichtlich unvernünftig. Sie können sich in jedem Fall in die Gefühle des Klägers und des Beklagten hineinversetzen, die sie dazu veranlassen, entweder eine kämpferische oder eine entgegenkommende Haltung einzunehmen. Doch langfristig sind die Abweichungen vom Erwartungswert wahrscheinlich kostspielig. Betrachten wir eine große Organisation, etwa die Stadt New York, und nehmen wir an, sie ist jedes Jahr mit 200 »mutwilligen« Prozessen konfrontiert, wobei jeder mit einer 5-prozentigen Wahrscheinlichkeit verknüpft ist, die Stadt 1 Million Dollar zu kosten. Nehmen wir weiterhin an, die Stadt könnte in jedem Fall gegen eine Zahlung von 100 000 Dollar die Einstellung des Verfahrens erreichen. Die Stadt erwägt zwei alternative Strategien, die sie auf alle derartigen Fälle anwenden wird: einen Vergleich schließen oder prozessieren. (Der Einfachheit halber lasse ich die Prozesskosten außer Betracht.)

– Wenn die Stadt in allen 200 Fällen prozessiert, wird sie zehn Verfahren verlieren, was einem Gesamtverlust von 10 Millionen Dollar entspricht.
– Wenn die Stadt in jedem Fall einen Vergleich über 100 000 Dollar schließt, wird ihr Gesamtverlust 20 Millionen Dollar betragen.

Wenn man über einen längeren Zeitraum viele ähnliche Entscheidungen betrachtet, erkennt man, dass es kostspielig ist, eine Prämie zu zahlen, um ein kleines Risiko eines hohen Verlusts zu vermeiden. Eine ähnliche Analyse gilt für jedes der Felder des viergeteilten Musters: Systematische Abweichungen vom Erwartungswert sind langfristig kostspielig – und diese Regel gilt sowohl für Risikoscheu wie für Risikofreude. Die konsequente Übergewichtung unwahrscheinlicher Ergebnisse – ein Merkmal intuitiver Entscheidungsfindung – führt schließlich zu schlechteren Ergebnissen.

## Zum Thema »Viergeteiltes Muster«

»Er neigt dazu, sich bei diesem mutwilligen Prozess auf einen Vergleich einzulassen, um einen außergewöhnlich hohen Verlust zu vermeiden. Das ist Übergewichtung geringer Wahrscheinlichkeiten. Da er wahrscheinlich mit vielen ähnlichen Problemen konfrontiert sein wird, wäre er gut beraten, nicht nachzugeben.«

»Wir haben uns bei unserem Urlaub nie auf ein Last-Minute-Schnäppchen verlassen. Wir sind bereit, uns die Sicherheit eine Menge kosten zu lassen.«

»Sie werden ihre Verluste so lange nicht verringern, wie noch eine Chance besteht, die Gewinnschwelle zu erreichen. Dies ist Risikofreude in der Verlustzone.«

»Sie wissen, dass das Risiko einer Gasexplosion sehr gering ist, aber sie wollen es abschwächen. Es ist ein Möglichkeitseffekt, und sie wollen ruhig schlafen können.«

# *30. Seltene Ereignisse*

Ich war mehrmals zu einer Zeit in Israel, als Selbstmordattentate in Bussen relativ häufig waren – auch wenn sie, absolut gesehen, recht selten gewesen sind. Zwischen Dezember 2001 und September 2004 gab es insgesamt 23 Bombenanschläge, die 236 Todesopfer forderten. Die Anzahl der täglichen Buspassagiere in Israel betrug damals etwa 1,3 Millionen. Für jeden einzelnen Passagier war das Risiko äußerst gering, aber die Bevölkerung empfand das anders. Die Menschen vermieden es, so gut sie konnten, Busse zu benutzen, und viele Passagiere verbrachten ihre Zeit im Bus damit, bei ihren Nachbarn nach Päckchen oder weiten Kleidungsstücken Ausschau zu halten, in denen sich eine Bombe verstecken mochte.

Ich hatte kaum Notwendigkeit, in Bussen zu fahren, da ich einen Mietwagen hatte, aber mit Verärgerung bemerkte ich, dass sich auch mein Verhalten änderte. Ich stellte fest, dass ich an einer roten Ampel immer mit einem flauen Gefühl neben einem Bus hielt, und ich fuhr schneller als gewöhnlich an, wenn die Ampel auf Grün sprang. Ich schämte mich über mich selbst, da ich es natürlich besser wusste. Ich wusste, dass das Risiko in Wirklichkeit zu vernachlässigen war und dass die kleinste Verhaltensänderung implizit einer winzigen Eintrittswahrscheinlichkeit ein ungewöhnlich hohes »Entscheidungsgewicht« zuschreiben würde. Tatsächlich war es wahrscheinlicher, bei einem Verkehrsunfall verletzt zu werden, als dadurch zu Schaden zu kommen, dass man neben einem Bus hielt, in dem sich ein Selbstmordattentäter in die Luft sprengte. Aber ich mied Busse nicht aus einer rationalen Sorge um mein Leben. Was mich antrieb, war die Erfahrung des Augenblicks: Wenn ich neben einem Bus stand, musste ich an Bomben denken, und diese Gedanken waren unangenehm. Ich mied Busse, weil ich an etwas anderes denken wollte.

Meine Erfahrung verdeutlicht, wie der Terrorismus funktioniert und weshalb er so wirkungsvoll ist: Er löst eine Verfügbarkeitskaskade aus. Ein äußerst anschauliches Bild von Tod und Verletzung, das ständig durch die Aufmerksamkeit der Medien und häufige Gespräche verstärkt wird, wird kognitiv in einem hohen Maße verfügbar, insbesondere wenn es mit einer bestimmten Situation wie etwa dem Anblick eines Busses verbunden ist.

Die emotionale Erregung ist assoziativ, automatisch und unkontrolliert, und sie erzeugt einen Impuls zu schützenden Handlungen. System 2 »weiß« vielleicht, dass die Wahrscheinlichkeit niedrig ist, aber dieses Wissen beseitigt nicht das selbsterzeugte Unbehagen und den Wunsch, es zu vermeiden.[1] System 1 kann nicht ausgeschaltet werden. Die Emotion steht nicht nur in einem deutlichen Missverhältnis zur Wahrscheinlichkeit, sie ist auch unempfindlich gegenüber dem genauen Grad der Wahrscheinlichkeit. Angenommen, zwei Städte sind gewarnt worden, dass sich in ihnen Selbstmordattentäter aufhalten. Die Bewohner der einen Stadt erfahren, dass zwei Attentäter bereit sind, zuzuschlagen. Die Bewohner einer anderen Stadt werden über die Anwesenheit eines Attentäters unterrichtet. Ihr Risiko ist um die Hälfte niedriger, aber fühlen sie sich deshalb viel sicherer?

Viele Geschäfte in New York City verkaufen Lotterielose, und die Nachfrage danach ist groß. Die Psychologie von Lotterien mit hohen Gewinnen ist vergleichbar mit der Psychologie des Terrorismus. Die prickelnde Möglichkeit, den Hauptpreis zu gewinnen, wird von der Gemeinschaft geteilt und durch Gespräche am Arbeitsplatz und zu Hause verstärkt. Der Kauf eines Loses wird sofort durch lustvolle Fantasien belohnt, so wie das Vermeiden eines Busses sofort durch Befreiung von Furcht belohnt wurde. In beiden Fällen ist die tatsächliche Wahrscheinlichkeit irrelevant; allein die Möglichkeit zählt. Die ursprüngliche Formulierung der Neuen Erwartungstheorie beinhaltete das Argument, dass »höchst unwahrscheinliche Ereignisse entweder ignoriert oder übergewichtet werden«, aber sie gab nicht die Bedingungen an, unter denen das eine oder das andere geschieht, und sie lieferte auch keine psychologische Interpretation. Meine gegenwärtige Sichtweise von Entscheidungsgewichten wurde stark beeinflusst von jüngeren Forschungen über die Rolle von Emotionen und Anschaulichkeit bei der Entscheidungsfindung.[2] Die Übergewichtung unwahrscheinlicher Ergebnisse wurzelt in Merkmalen von System 1, die uns mittlerweile vertraut sind. Emotionalität und Anschaulichkeit beeinflussen Abrufflüssigkeit, Verfügbarkeit und Wahrscheinlichkeitsurteile – und erklären daher unsere überzogene Reaktion auf die wenigen seltenen Ereignisse, die wir nicht ignorieren können.

## Überschätzen und Übergewichten

> Wie hoch schätzen Sie die Wahrscheinlichkeit ein, dass der nächste Präsident der Vereinigten Staaten der Kandidat einer dritten Partei sein wird?

> Wie viel bezahlen Sie für eine Wette, bei der Sie 1000 Dollar erhalten, wenn der nächste Präsident der Vereinigten Staaten der Kandidat einer dritten Partei ist, und nichts, wenn es sich anders verhält?

Die beiden Fragen sind unterschiedlich, aber offensichtlich miteinander verwandt. Die erste fordert Sie auf, die Wahrscheinlichkeit eines unwahrscheinlichen Ereignisses zu beurteilen. Die zweite bittet Sie, das gleiche Ereignis mit einem Entscheidungsgewicht zu versehen, indem Sie eine Wette darauf abschließen.

Wie treffen Menschen Urteile, und wie schreiben sie Entscheidungsgewichte zu? Wir beginnen mit zwei einfachen Antworten, die wir anschließend genauer bestimmen. Hier sind die grob vereinfachten Antworten:

- Menschen überschätzen die Wahrscheinlichkeiten unwahrscheinlicher Ereignisse.
- Menschen übergewichten unwahrscheinliche Ereignisse bei ihren Entscheidungen.

Obgleich Überschätzung und Übergewichtung verschiedene Phänomene sind, sind an beiden dieselben psychologischen Mechanismen beteiligt: fokussierte Aufmerksamkeit, Bestätigungsfehler und kognitive Leichtigkeit.

Konkrete Beschreibungen lösen die assoziative Maschinerie von System 1 aus. Als Sie über den unwahrscheinlichen Sieg eines dritten Kandidaten

nachdachten, arbeitete Ihr assoziatives System in seiner üblichen bestätigenden Weise, wobei es selektiv Informationen, Beispiele und Bilder abrief, die die Aussage wahr machen würden. Der Prozess war verzerrt, aber er war keine Übung der Fantasie. Sie suchten nach einem plausiblen Szenario, das sich den Zwängen der Wirklichkeit fügt; Sie haben sich nicht einfach ausgemalt, wie eine Märchenfee den Kandidaten einer dritten Partei als Präsidenten einsetzt. Ihr Wahrscheinlichkeitsurteil wurde letztlich von der kognitiven Leichtigkeit beziehungsweise Abrufflüssigkeit bestimmt, mit der Ihnen ein plausibles Szenario einfiel.

Sie konzentrieren sich nicht immer auf das Ereignis, das Sie einschätzen sollen. Wenn das Zielereignis sehr wahrscheinlich ist, konzentrieren Sie sich auf seine Alternative. Betrachten Sie dieses Beispiel:

> Wie wahrscheinlich ist es, dass ein Neugeborenes, das in Ihrem örtlichen Krankenhaus zur Welt kam, innerhalb von drei Tagen mit seiner Mutter entlassen wird?

Sie sollten die Wahrscheinlichkeit beurteilen, mit der das Neugeborene aus dem Krankenhaus entlassen wird, aber Sie haben sich höchstwahrscheinlich auf die Ereignisse konzentriert, die dazu führen könnten, dass ein Baby *nicht* innerhalb der üblichen Zeit entlassen wird. Unser Intellekt besitzt die nützliche Fähigkeit, sich spontan auf alles zu konzentrieren, was seltsam, fremdartig oder ungewöhnlich ist. Sie haben schnell verstanden, dass Neugeborene in den Vereinigten Staaten (nicht alle Staaten haben die gleichen Standards) normalerweise innerhalb von zwei oder drei Tagen nach ihrer Geburt nach Hause entlassen werden, sodass sich Ihre Aufmerksamkeit auf die ungewöhnliche Alternative gerichtet hat. Das unwahrscheinliche Ereignis rückte ins Zentrum der Aufmerksamkeit. Wahrscheinlich wird die Verfügbarkeitsheuristik aktiviert: Ihr Urteil wurde vermutlich von der Anzahl der Szenarien medizinischer Probleme, die Ihnen einfielen, und von der Leichtigkeit, mit der sie Ihnen einfielen, bestimmt. Weil Sie im Bestätigungsmodus waren, war Ihre Schätzung der Häufigkeit von Problemen vermutlich zu hoch.

Die Wahrscheinlichkeit eines seltenen Ereignisses wird wahrscheinlich überschätzt, wenn die Alternative nicht ganz genau angegeben wird. Mein

Lieblingsbeispiel stammt aus einer Studie, die der Psychologe Craig Fox durchführte, als er bei Amos studierte.[3] Für diese Studie rekrutierte Fox Fans professioneller Basketballmannschaften; er bat sie um etliche Urteile und Entscheidungen bezüglich des Gewinners der NBA-Play-offs. Sie sollten insbesondere die Wahrscheinlichkeit einschätzen, mit der jedes der acht teilnehmenden Teams das Play-off gewinnt; der Sieg jedes Teams war seinerseits das fokale Ereignis.

Sie ahnen sicher schon, was herauskam, aber die Stärke des Effekts, die Fox beobachtete, überrascht Sie vielleicht. Stellen Sie sich einen Fan vor, der die Wahrscheinlichkeit einschätzen soll, dass die Chicago Bulls das Turnier gewinnen werden. Das fokale Ereignis ist gut definiert, aber seine Alternative – eines der anderen sieben Teams trägt den Sieg davon – ist diffus und weniger plastisch. Gedächtnis und Fantasie des Fans, die im Bestätigungsmodus arbeiten, versuchen, einen Sieg für die Bulls zu konstruieren. Wenn dieselbe Person als Nächstes die Chancen der Lakers beurteilen soll, wird die gleiche selektive Aktivierung zugunsten dieses Teams arbeiten. Die acht besten Profi-Basketballteams in den Vereinigten Staaten sind alle sehr gut, und man kann sich durchaus vorstellen, dass ein relativ schwaches Team unter ihnen als Sieger hervorgeht. Das Ergebnis: Die für die acht Mannschaften nacheinander generierten Wahrscheinlichkeitsurteile summierten sich auf 240 Prozent! Das ist natürlich absurd, weil sich die Summe der Wahrscheinlichkeiten der acht Ereignisse auf 100 Prozent addieren *muss*. Die Absurdität verschwand, als dieselben Beurteiler gefragt wurden, ob der Sieger aus der Eastern oder der Western Conference kommen würde. Das fokale Ereignis und seine Alternativen waren in dieser Frage in gleicher Weise spezifiziert, und die Wahrscheinlichkeitsurteile summierten sich auf 100 Prozent.[4]

Um Entscheidungsgewichte zu beurteilen, bat Fox die Basketballfans auch, auf das Turnierergebnis zu wetten. Sie schrieben jeder Wette ein Bargeldäquivalent zu (ein Bargeldbetrag, der genauso attraktiv war wie das Spielen der Wette). Das Gewinnen der Wette wäre mit einer Auszahlung von 160 Dollar verbunden. Die Summe der Bargeldäquivalente für die acht Mannschaften betrug 287 Dollar. Ein Teilnehmer, der alle acht Wetten einging, würde im Schnitt einen Verlust von 127 Dollar machen! Die Teilnehmer wussten genau, dass an dem Turnier acht Mannschaften teilnahmen und dass, wenn man auf sie alle setzte, die durchschnittliche

Auszahlung 160 Dollar nicht übersteigen konnte, aber sie übergewichteten trotzdem. Die Fans überschätzten nicht nur die Wahrscheinlichkeit der Ereignisse, auf die sie sich konzentrierten – sie waren auch allzu bereit, auf diese zu wetten.

Diese Ergebnisse warfen auch ein neues Licht auf den Planungsfehler und andere Manifestationen von Optimismus. Die erfolgreiche Ausführung eines Plans ist konkret und leicht vorstellbar, wenn man versucht, das Ergebnis eines Projekts vorherzusagen. Dagegen ist die Alternative des Scheiterns diffus, weil Dinge auf zahllose Weise schiefgehen können. Unternehmer und Investoren, die ihre Gewinnaussichten beurteilen, neigen dazu, ihre Chancen zu überschätzen und ihre Schätzungen überzugewichten.

## Anschauliche Ergebnisse

Wie wir gesehen haben, unterscheidet sich die Neue Erwartungstheorie von der Nutzentheorie in der Beziehung, die sie zwischen Wahrscheinlichkeit und Entscheidungsgewicht herstellt. In der Nutzentheorie sind Entscheidungsgewichte und Wahrscheinlichkeiten identisch. Das Entscheidungsgewicht eines sicheren Ereignisses ist 100, und das Gewicht, das einer 90-prozentigen Chance entspricht, ist genau 90, was exakt das Neunfache des Entscheidungsgewichts für eine 10-prozentige Chance entspricht. In der Neuen Erwartungstheorie wirken sich Schwankungen der Wahrscheinlichkeit weniger stark auf die Entscheidungsgewichte aus. Ein Experiment, das ich weiter vorn erwähnte, gelangte zu dem Ergebnis, dass das Entscheidungsgewicht für eine 90-prozentige Chance 71,2 und für eine 10-prozentige Chance 18,6 ist. Das Verhältnis der Wahrscheinlichkeiten war 9,0, aber das Verhältnis der Entscheidungsgewichte war nur 3,83, was auf eine ungenügende Wahrscheinlichkeitsempfindlichkeit in diesem Bereich hindeutet. In beiden Theorien hängen die Entscheidungsgewichte nur von der Wahrscheinlichkeit ab, nicht von dem Ergebnis. Beide Theorien sagen vorher, dass die Entscheidungsgewichte für eine 90-prozentige Chance die gleichen sind, egal ob es darum geht, 100 Dollar zu gewinnen, ein Dutzend Rosen zu erhalten oder einen Elektroschock zu bekommen.[5] Diese theoretische Vorhersage erweist sich als falsch.

Psychologen an der Universität von Chicago veröffentlichten einen Aufsatz mit dem reizvollen Titel »Money, Kisses, and Electric Shocks: On the Affective Psychology of Risk« (»Geld, Küsse und Elektroschocks: Über die Affektpsychologie von Risiken«). Sie fanden heraus, dass die Bewertung von Lotterien viel weniger wahrscheinlichkeitssensitiv war, wenn die (fiktiven) Ergebnisse emotionaler Natur waren (»Ihren liebsten Filmstar treffen und küssen« oder »einen schmerzhaften, aber ungefährlichen Elektroschock bekommen«), als wenn es sich bei den Ergebnissen um Geldgewinne oder -verluste handelte. Dies war kein isolierter Befund. Andere Forscher hatten mithilfe physiologischer Messgrößen wie dem Herzschlag herausgefunden, dass die Furcht vor einem unmittelbar bevorstehenden Elektroschock in keinem nennenswerten Zusammenhang mit der Wahrscheinlichkeit stand, den Schock zu erhalten.

Die bloße Möglichkeit eines Schocks löste eine voll ausgebildete Furchtreaktion aus. Die Forschergruppe aus Chicago erklärte dies damit, dass »affektgeladene Vorstellungsbilder« eine an der Wahrscheinlichkeit orientierte Reaktion überwältigt hätten. Zehn Jahre später stellte ein Team von Psychologen von der Universität Princeton diese Schlussfolgerung infrage. Die Forschergruppe aus Princeton behauptete, die beobachtete geringe Wahrscheinlichkeitsempfindlichkeit für emotionale Ergebnisse sei normal. Glücksspiele um Geld seien die Ausnahme. Die Wahrscheinlichkeitsempfindlichkeit ist für diese Glücksspiele relativ hoch, weil sie einen bestimmten Erwartungswert haben.

> Welcher Geldbetrag ist genauso attraktiv wie jede dieser Lotterien?
>
> a. 84-prozentige Chance, 59 Dollar zu gewinnen.
> b. 84-prozentige Chance, ein Dutzend rote Rosen in einer Glasvase zu bekommen.

Was fällt Ihnen auf? Der markante Unterschied besteht darin, dass Frage A viel leichter zu beantworten ist als Frage B. Sie haben nicht aufgehört, den Erwartungswert der Wette zu berechnen, aber Sie wussten vermutlich sehr schnell, dass er nicht weit von 50 Dollar entfernt ist (tatsächlich beträgt er 49,56 Dollar), und die vage Schätzung genügte, um einen hilfreichen Anker bereitzustellen, als Sie sich über die Höhe eines gleich attraktiven Geldgeschenks Gedanken machten. Für Frage B ist kein solcher Anker verfügbar, und sie ist daher viel schwerer zu beantworten. Die Befragten beurteilten auch die Bargeldäquivalente von Lotterien mit einer 21-prozentigen Chance, die beiden Ergebnisse zu gewinnen. Erwartungsgemäß war die Differenz zwischen den Lotterien mit hoher und niedriger Gewinnwahrscheinlichkeit für Geld viel ausgeprägter als für Rosen.

Um sein Argument, dass die Unempfindlichkeit für die Wahrscheinlichkeit nicht durch Emotionen verursacht wird, zu untermauern, verglich das Team aus Princeton die Bereitschaft, zu zahlen, um Lotterien zu vermeiden:

21-prozentige Chance (oder 84-prozentige Chance), ein Wochenende damit zu verbringen, die Drei-Zimmer-Wohnung eines anderen zu streichen.

21-prozentige Chance (oder 84-prozentige Chance), drei Toilettenkabinen in einem Studentenwohnheim zu reinigen, nachdem sie übers Wochenende benutzt wurden.

Das zweite Ergebnis ist sicherlich emotional viel stärker aufgeladen als das erste, aber die Entscheidungsgewichte für die beiden Ergebnisse unterschieden sich nicht. Offensichtlich ist die Stärke der Emotion nicht die Antwort.

Ein anderes Experiment erbrachte ein überraschendes Ergebnis. Die Teilnehmer erhielten explizite Preisinformationen zusammen mit der verbalen Beschreibung des Gewinns. Ein Beispiel könnte sein:

84-prozentige Gewinnchance für ein Dutzend rote Rosen in einer Glasvase. Wert: 59 Dollar.

21-prozentige Gewinnchance für ein Dutzend rote Rosen in einer Glasvase. Wert: 59 Dollar.

Der monetäre Erwartungswert dieser Lotterien lässt sich leicht errechnen, aber die Hinzufügung eines bestimmten Geldwertes änderte nichts an den Ergebnissen: Selbst unter dieser Bedingung blieben die Bewertungen unempfindlich gegenüber der Wahrscheinlichkeit. Menschen, die in dem Geschenk eine Chance sahen, Rosen zu bekommen, benutzten bei der Bewertung des Glücksspiels die Preisinformationen nicht als Anker. Wie

Wissenschaftler manchmal sagen, ist dies ein überraschender Befund, der uns etwas sagen will. Aber was will er uns sagen?

Er will uns meines Erachtens sagen, dass eine konkrete und anschauliche Darstellung des Ergebnisses, unabhängig davon, ob es emotional aufgeladen ist oder nicht, die Bedeutung der Wahrscheinlichkeit bei der Beurteilung einer unsicheren Aussicht reduziert. Diese Hypothese legt eine Vorhersage nahe, von der ich hinlänglich überzeugt bin: Die Berechnung wird auch gestört, wenn ich ein monetäres Ergebnis mit belanglosen, aber anschaulichen Details anreichere. Vergleichen Sie Ihre Bargeldäquivalente für die folgenden Ergebnisse:

> 21-prozentige (oder 84-prozentige) Chance, nächsten Montag 59 Dollar zu bekommen.

> 21-prozentige (oder 84-prozentige) Chance, nächsten Montagmorgen einen großen blauen Pappumschlag mit 59 Dollar darin zu bekommen.

Die neue Hypothese lautet, dass im zweiten Fall die Wahrscheinlichkeitsempfindlichkeit geringer ist, weil der blaue Umschlag eine reichhaltigere und flüssigere Repräsentation hervorruft als der abstrakte Begriff einer Geldsumme. Sie haben dieses Ereignis in Ihrem Geist konstruiert, und das anschauliche Bild des Ergebnisses existiert dort, auch wenn Sie wissen, dass seine Wahrscheinlichkeit gering ist. Auch die kognitive Leichtigkeit trägt zum Sicherheitseffekt bei: Wenn man ein anschauliches Bild eines Ereignisses im Sinn hat, ist die Möglichkeit, dass es nicht eintritt, ebenfalls anschaulich repräsentiert und übergewichtet. Die Kombination eines verstärkten Möglichkeitseffekts mit einem verstärkten Sicherheitseffekt lässt wenig Raum dafür, dass Entscheidungsgewichte von den Wahrscheinlichkeiten von 21 und 84 Prozent abweichen.

## Anschauliche Wahrscheinlichkeiten

Die Annahme, dass Abrufleichtigkeit, Anschaulichkeit und Leichtigkeit der Imagination zu den Entscheidungsgewichten beitragen, wird von vielen anderen Beobachtungen gestützt. Die Teilnehmer an einem bekannten Experiment sollen eine Murmel aus einer von zwei Urnen ziehen, wobei rote Murmeln einen Preis gewinnen:

> Urne A enthält zehn Murmeln, von denen eine rot ist.
> Urne B enthält hundert Murmeln, von denen acht rot sind.

Für welche Urne würden Sie sich entscheiden? Die Gewinnchancen liegen bei 10 Prozent in Urne A und 8 Prozent in Urne B, sodass es leicht sein sollte, die richtige Wahl zu treffen. Aber das ist nicht der Fall: Etwa 30 bis 40 Prozent der Studenten wählen die Urne mit der größeren *Anzahl* gewinnbringender Murmeln statt die Urne, die mit einer höheren Gewinnchance verbunden ist. Seymour Epstein hat behauptet, die Ergebnisse verdeutlichten die typische oberflächliche Verarbeitungsweise von System 1 (das er das »erfahrungsbasierte System« nennt).[6]

Die bemerkenswert unvernünftigen Entscheidungen, die Menschen in dieser Situation treffen, haben erwartungsgemäß die Aufmerksamkeit vieler Forscher auf sich gezogen. Dieser Verzerrung wurden verschiedene Namen gegeben; ich folge Paul Slovic und nenne sie »Nenner-Vernachlässigung« *(denominator neglect)*. Wenn Ihre Aufmerksamkeit auf die gewinnträchtigen Murmeln gerichtet ist, berücksichtigen Sie nicht mit der gleichen Sorgfalt die Anzahl der nicht gewinnenden Murmeln. Anschauliche Vorstellungsbilder tragen zur Vernachlässigung des Nenners bei, zumindest erlebe ich es so. Wenn ich mir die kleine Urne vorstelle, sehe ich eine einzelne rote Murmel auf einem vage definierten Untergrund weißer Murmeln. Wenn ich mir die größere Urne vorstelle, sehe ich acht gewinnbringende rote Murmeln vor einem verschwommenen Hintergrund weißer Murmeln, was ein hoffnungsvolleres Gefühl auslöst. Die markante

Anschaulichkeit der gewinnbringenden Murmeln erhöht das Entscheidungsgewicht dieses Ereignisses und verstärkt so den Möglichkeitseffekt. Natürlich gilt das Gleiche für den Sicherheitseffekt. Wenn ich eine 90-prozentige Gewinnchance habe, ist das Ereignis, nicht zu gewinnen, hervorstechender, wenn zehn von hundert Murmeln »Verlierer« sind, als wenn eine von zehn Murmeln zum gleichen Ergebnis führt.

Das Konzept der Nenner-Vernachlässigung hilft zu erklären, weshalb verschiedene Weisen, Risiken zu kommunizieren, so unterschiedliche Wirkungen entfalten. Sie lesen, dass »ein Impfstoff, der Kinder vor einer tödlichen Erkrankung schützt, mit dem 0,001-prozentigen Risiko einer dauerhaften Behinderung verbunden ist«. Das Risiko scheint klein zu sein. Betrachten Sie jetzt eine andere Beschreibung des gleichen Risikos: »Eines von 100 000 geimpften Kindern wird dauerhaft behindert sein.« Die zweite Aussage macht etwas mit Ihnen, was die erste nicht tut: Sie ruft ein Vorstellungsbild eines einzelnen Kindes auf, das durch eine Impfung dauerhaft behindert ist; die 99 999 ohne gesundheitliche Schäden geimpften Kinder sind in den Hintergrund verschwunden. Wie von der Nenner-Vernachlässigung vorhergesagt, werden Ereignisse mit niedriger Wahrscheinlichkeit viel stärker gewichtet, wenn sie in Kategorien von relativen Häufigkeiten (Wie viele?) beschrieben werden, als wenn sie in abstrakteren Begriffen von »Chancen«, »Risiko« oder »Wahrscheinlichkeit« (Wie wahrscheinlich?) ausgedrückt werden. Wie wir gesehen haben, kommt System 1 viel besser mit Individuen als mit Kategorien zurecht.

Das Häufigkeitsformat hat einen starken Effekt. In einer Studie stuften die Teilnehmer, die von einer »Krankheit [erfuhren], die 1286 von 10 000 Menschen umbringt«, diese als gefährlicher ein als eine »Krankheit, die 24,14 Prozent der von ihr Betroffenen das Leben kostet«.[7] Die erste Krankheit scheint bedrohlicher zu sein als die zweite, obgleich das Risiko im ersten Fall nur halb so groß ist wie im zweiten! In einer noch direkteren Demonstration der Nenner-Vernachlässigung wurde »eine Krankheit, die bei 1286 von 10 000 Erkrankten tödlich verläuft«, für gefährlicher gehalten als eine Krankheit, »die bei 24,4 von 100 Erkrankten tödlich verläuft«. Der Effekt würde mit Sicherheit verringert oder beseitigt, wenn Teilnehmer zu einem direkten Vergleich der beiden Formulierungen aufgefordert würden, eine Aufgabe, die sich explizit an System 2 wendet. Das Leben ist jedoch für gewöhnlich ein Between-Subjects-Experiment, bei dem man zu einem bestimmten Zeitpunkt immer nur eine Formulierung sieht. Es bedürfte eines

außergewöhnlich aktiven Systems 2, um alternative Formulierungen derjenigen zu erzeugen, die Sie gesehen haben, und zu bemerken, dass sie eine andere Reaktion hervorrufen.

Selbst erfahrene Rechtspsychologen und -psychiater sind nicht immun gegen die Wirkungen des Formats, in dem Risiken ausgedrückt werden.[8] In einem Experiment sollten Fachleute beurteilen, ob es ungefährlich wäre, einen Patienten, Mr. Jones, der schon etliche Gewalttaten begangen hatte, aus einer psychiatrischen Klinik zu entlassen. Die Informationen, die sie erhielten, beinhalteten ein Sachverständigengutachten mit der Risikoeinschätzung. Derselbe statistische Sachverhalt wurde auf zwei verschiedene Weisen beschrieben:

> Ähnliche Patienten wie Mr Jones werden mit einer geschätzten Wahrscheinlichkeit von 10 Prozent in den ersten Monaten nach ihrer Entlassung eine Gewalttat begehen.

> Von je hundert Patienten, die mit Mr Jones vergleichbar sind, werden schätzungsweise zehn in den ersten Monaten nach ihrer Entlassung eine Gewalttat gegen andere verüben.

Die Fachleute, die das Häufigkeitsformat sahen, lehnten die Entlassung fast doppelt so oft ab (41 Prozent im Vergleich zu 21 Prozent im Wahrscheinlichkeitsformat). Die anschaulichere Beschreibung erzeugt ein höheres Entscheidungsgewicht für dieselbe Wahrscheinlichkeit.

Die Macht des Formates schafft Gelegenheiten zur Manipulation, die Menschen, die auf ihren Vorteil bedacht sind, auszunutzen wissen. Slovic und seine Kollegen zitieren einen Artikel, in dem es heißt, »jährlich werden landesweit etwa tausend Morde von psychisch schwer kranken Menschen begangen, die ihre Medikamente nicht einnehmen«. Eine andere Formulierung derselben Tatsache lautet: »Von 273 000 000 Amerikanern werden in diesem Jahr tausend auf diese Weise ums Leben kommen.« Eine weitere Variante lautet: »Die Wahrscheinlichkeit, durch eine solche Person umgebracht zu werden, beträgt, auf ein Jahr bezogen, etwa 0,00036

Prozent.« Noch eine andere lautet: »Jedes Jahr werden tausend Amerikaner auf diese Weise umkommen, weniger als ein Dreißigstel der Anzahl der Menschen, die durch Suizid sterben, und etwa ein Viertel der Anzahl der Menschen, die an Kehlkopfkrebs sterben.« Slovic weist darauf hin, dass »diese interessierten Personen aus ihrer Motivation keinen Hehl machen: sie *wollen* der Bevölkerung mit der Gewaltneigung psychisch kranker Menschen einen Schrecken einjagen, in der Hoffnung, dass sich diese Furcht in einer besseren Finanzierung psychiatrischer Betreuungseinrichtungen niederschlägt«.

Ein guter Anwalt, der einen DNA-Beweis in Zweifel ziehen will, sagt vor den Geschworenen nicht, dass »die Wahrscheinlichkeit einer falschen Übereinstimmung 0,1 Prozent beträgt«. Die Aussagen, dass »eine falsche Übereinstimmung bei einem von tausend Fällen auftritt«, wird die Schwelle begründeter Zweifel viel eher überwinden.[9] Die Geschworenen, die diese Worte hören, werden aufgefordert, das Bild des Mannes zu erzeugen, der vor ihnen im Gerichtssaal sitzt und wegen eines fehlerhaften DNA-Beweises verurteilt wird. Der Staatsanwalt wird selbstverständlich die abstraktere Formulierung vorziehen – in der Hoffnung, die Köpfe der Juroren mit Dezimalpunkten zu füllen.

## Entscheidungen auf der Basis globaler Eindrücke

Die Studiendaten legen die Hypothese nahe, dass die fokale Aufmerksamkeit und die Auffälligkeit (Salienz) sowohl zur Überschätzung unwahrscheinlicher Ereignisse als auch zur Übergewichtung unwahrscheinlicher Ergebnisse beitragen. Die Auffälligkeit wird durch die bloße Erwähnung eines Ereignisses, seine Anschaulichkeit und das Format, in dem die Wahrscheinlichkeit beschrieben wird, verstärkt. Es gibt natürlich Ausnahmen, bei denen die Fokussierung auf ein Ereignis seine Wahrscheinlichkeit nicht erhöht: Fälle, in denen eine fehlerhafte Theorie ein Ereignis auch dann als unmöglich erscheinen lässt, wenn man darüber nachdenkt, oder Fälle, in denen einen die Unfähigkeit, sich vorzustellen, wie ein Ergebnis zustande kommen könnte, in der Überzeugung wiegt, dass es sich nicht ereignen wird. Die Neigung zur Überschätzung und Übergewichtung auffälliger Ereignisse ist keine absolute Regel, aber sie ist ausgeprägt und robust.

In den letzten Jahren interessierte man sich sehr für Studien über Wahlen auf Erfahrungsgrundlage, die anderen Regeln folgen als Wahlen auf Beschreibungsgrundlage, die in der Neuen Erwartungstheorie analysiert werden.[10] Teilnehmer an einem typischen Experiment haben zwei Tasten vor sich. Jede Taste gibt, wenn man sie drückt, eine Geldbelohnung oder nichts frei, und das Ergebnis kommt zufällig zustande, laut den Spezifikationen einer Lotterie (zum Beispiel »eine 5-prozentige Chance, 12 Dollar zu gewinnen«, oder »eine 95-prozentige Chance, 1 Dollar zu gewinnen«). Der Prozess ist wirklich zufallsabhängig, sodass es keine Garantie dafür gibt, dass eine Stichprobe, die ein Teilnehmer sieht, exakt den statistischen Randbedingungen entspricht. Die mit den beiden Tasten verbundenen Erwartungswerte sind ungefähr gleich, aber einer ist riskanter (variabler) als der andere. (So mag zum Beispiel das Drücken einer Taste in 5 Prozent der Versuchsdurchgänge 10 Dollar erbringen, während das Drücken der anderen in 50 Prozent der Durchgänge 1 Dollar erbringt). Bei Wahlen auf Erfahrungsgrundlage durchläuft der Proband viele Versuche, in denen er die Folgen des Drückens der einen oder der anderen Taste beobachten kann. Bei dem entscheidenden Durchgang wählt er eine der beiden Tasten und bekommt das Ergebnis dieses Versuchs. Bei einer Wahl

auf Beschreibungsgrundlage wird dem Probanden die riskante Gewinnaussicht, die mit jeder Taste verbunden ist (wie etwa »mit 5-prozentiger Chance 12 Dollar gewinnen«) gezeigt, und er wird aufgefordert, eine auszuwählen. Wie nach der Neuen Erwartungstheorie zu erwarten, geht die Wahl auf Beschreibungsgrundlage mit einem Möglichkeitseffekt einher – seltene Ergebnisse werden im Verhältnis zu ihrer Wahrscheinlichkeit übergewichtet. In scharfem Gegensatz dazu wird die Übergewichtung bei erfahrungsbasierten Wahlen nicht beobachtet, und die Untergewichtung ist weit verbreitet.

Die experimentelle Situation von Wahlen auf Erfahrungsgrundlage soll viele Situationen repräsentieren, in denen wir variablen Ergebnissen von derselben Quelle ausgesetzt sind. Ein Restaurant, das für gewöhnlich gut ist, mag hin und wieder ein vorzügliches oder ein miserables Gericht servieren. Ihr Freund ist in der Regel gute Gesellschaft, aber manchmal wird er launisch und aggressiv. Kalifornien ist eine erdbebengefährdete Region, tatsächlich kommt es dort jedoch nur selten zu Erdbeben. Die Ergebnisse vieler Experimente deuten darauf hin, dass seltene Ereignisse nicht übergewichtet sind, wenn wir Entscheidungen treffen wie etwa die Wahl eines Restaurants oder das Sichern des Heizkessels, um Erdbebenschäden zu verringern.

Die Interpretation von Wahlen auf Erfahrungsgrundlage ist noch nicht abgeschlossen, aber es besteht allgemeine Einigkeit über eine Hauptursache der Untergewichtung seltener Ereignisse sowohl bei Experimenten als auch in der Wirklichkeit: Viele Teilnehmer erleben das seltene Ereignis nicht![11] Die meisten Kalifornier haben nie ein größeres Erdbeben erlebt, und bis 2007 hatte kein Banker persönlich eine verheerende Finanzkrise durchgemacht. Ralph Hertwig und Ido Erev schreiben: »Wahrscheinlichkeiten seltener Ereignisse (wie etwa das Platzen von Immobilienblasen) werden nicht so stark beachtet, wie sie es nach ihren objektiven Wahrscheinlichkeiten verdienen würden.«[12] Als Beispiel führen sie die verhaltene Reaktion der Bevölkerung auf langfristige Umweltgefahren an. Diese Beispiele der Vernachlässigung sind bedeutsam und leicht erklärbar, aber es kommt auch dann zu Untergewichtung, wenn Menschen das seltene Ereignis selbst erlebt haben. Angenommen, Sie haben eine komplizierte Frage, die zwei Kollegen auf derselben Etage wahrscheinlich beantworten könnten. Sie kennen sie seit Jahren und hatten

viele Gelegenheiten, sie zu beobachten und ihren Charakter zu erleben. Adele ist ziemlich beständig und im Allgemeinen hilfsbereit, wenn auch nicht in einem außerordentlichen Maße. Brian ist in der Regel nicht ganz so freundlich und hilfsbereit wie Adele, aber manchmal ist er äußerst großzügig mit seiner Zeit und seinem Rat. An wen wenden Sie sich?

Betrachten Sie zwei mögliche Sichtweisen dieser Entscheidung:

– Es ist eine Wahl zwischen zwei Glücksspielen. Adele ist eher eine sichere Sache; die Option Brian wird eher ein leicht schlechteres Ergebnis bringen, mit einer geringen Wahrscheinlichkeit eines sehr guten Ergebnisses. Das seltene Ereignis wird durch einen Möglichkeitseffekt übergewichtet, der Brian begünstigt.
– Es ist eine Wahl zwischen Ihren globalen Eindrücken von Adele und Brian. Die positiven und die negativen Erfahrungen, die Sie hatten, werden in Ihrer Repräsentation ihres normalen Verhaltens zusammengefasst. Sofern das seltene Ereignis nicht so extrem ist, dass es Ihnen gesondert einfällt (Brian hat einmal einen Kollegen, der ihn um Hilfe bat, übel beschimpft), wird die Norm zugunsten typischer und neuerer Erfahrungen verzerrt sein und Adele begünstigen.

In einem Zwei-Systeme-Intellekt erscheint die zweite Interpretation weitaus plausibler. System 1 erzeugt globale Repräsentationen von Adele und Brian, zu denen eine emotionale Einstellung und eine Tendenz zur Annäherung oder Vermeidung gehört. Man muss lediglich diese Tendenzen vergleichen, um herauszufinden, an welche Tür Sie klopfen werden. Wenn Ihnen das seltene Ereignis nicht explizit einfällt, wird es nicht übergewichtet werden. Dieselbe Idee lässt sich leicht auf Experimente über Wahlen auf Erfahrungsgrundlage anwenden. In dem Maße, wie die beiden Tasten im Zeitablauf Ergebnisse liefern, entwickeln sie integrierte »Persönlichkeiten«, die mit emotionalen Reaktionen verknüpft werden.

Die Bedingungen, unter denen seltene Ereignisse ignoriert oder übergewichtet werden, lassen sich heute besser verstehen als zu der Zeit, als die Neue Erwartungstheorie formuliert wurde. Die Wahrscheinlichkeit eines seltenen Ereignisses wird (oft, nicht immer) aufgrund des Bestätigungsfehlers des Gedächtnisses überschätzt. Wenn Sie über dieses Ereignis nachdenken, versuchen Sie, es in Ihrem Geist als wahr anzunehmen. Ein seltenes Ereignis wird übergewichtet, wenn es gesondert Aufmerksamkeit auf sich zieht. Gesonderte Aufmerksamkeit ist faktisch

garantiert, wenn Chancen explizit beschrieben werden (»eine 99-prozentige Chance, 1000 Dollar zu gewinnen, und eine 1-prozentige Chance, nichts zu gewinnen«). Zwanghafte Ängste (der Bus in Jerusalem), plastische Bilder (die Rosen), konkrete Repräsentationen (1 von 1000) und explizite Erinnerungen (wie bei Wahlen auf der Basis von Beschreibungen) tragen alle zur Übergewichtung bei. Und wenn keine Übergewichtung stattfindet, kommt es zu Vernachlässigung. Unser Gehirn ist nicht dafür ausgelegt, geringe Wahrscheinlichkeiten richtig einzuschätzen. Für die Bewohner eines Planeten, dem möglicherweise Ereignisse bevorstehen, die bislang noch niemand erlebt hat, sind dies keine guten Neuigkeiten.

## Zum Thema »Seltene Ereignisse«

»Tsunamis sind selbst in Japan sehr selten, aber das Bild ist so anschaulich und eindringlich, dass Touristen zwangsläufig ihre Wahrscheinlichkeit überschätzen.«

»Es ist der vertraute Katastrophenzyklus. Es beginnt mit Übertreibung und Übergewichtung, dann setzt Vernachlässigung ein.«

»Wir sollten uns nicht auf ein einzelnes Szenario konzentrieren, oder wir werden seine Wahrscheinlichkeit überschätzen. Wir wollen spezifische Alternativen angeben und dafür sorgen, dass sich die Wahrscheinlichkeiten auf 100 Prozent addieren.«

»Sie wollen, dass die Menschen wegen des Risikos beunruhigt sind. Aus diesem Grund beschreiben sie es als ›1 Todesfall pro 1000‹. Sie setzen auf die Vernachlässigung des Nenners.«

# *31. Risikostrategien*

Stellen Sie sich vor, Sie sind mit dem folgenden Paar konkurrierender Entscheidungen konfrontiert. Prüfen Sie zunächst beide Entscheidungen, bevor Sie Ihre Wahl treffen.

Entscheidung 1: Wählen Sie zwischen:

A Sicherer Gewinn von 240 Dollar.
B 25-prozentige Chance, 1000 Dollar zu gewinnen, und 75-prozentige Chance, nichts zu gewinnen.

Entscheidung 2: Wählen Sie zwischen:

C Sicherer Verlust von 750 Dollar.
D 75-prozentige Chance, 1000 Dollar zu verlieren, und 25-prozentige Chance, nichts zu verlieren.

Dieses Paar von Wahlproblemen hat einen wichtigen Platz in der Geschichte der Neuen Erwartungstheorie, und es gibt uns neue Aufschlüsse über rationales Entscheidungsverhalten. Als Sie die beiden Probleme überflogen, war Ihre anfängliche Reaktion auf die sicheren Optionen (A und C) Hingezogenheit zur ersten und Abneigung gegenüber der zweiten. Die emotionale Beurteilung eines »sicheren Gewinns« und eines »sicheren Verlusts« ist eine automatische Reaktion von System 1, die zweifellos vor der anstrengenderen (und optionalen) Berechnung der Erwartungswerte beider Lotterien (ein Gewinn von 250 Dollar beziehungsweise ein Verlust von 750 Dollar) geschieht. Die Wahlen der meisten Menschen entsprechen den Vorlieben von System 1, und eine große Mehrheit zieht A Option B vor und D Option C. Wie bei vielen anderen Wahlen, bei denen es um mittlere oder hohe Wahrscheinlichkeiten geht, neigen Menschen dazu, im Bereich von Gewinnen risikoscheu und im Bereich von Verlusten risikofreudig zu

sein. In dem ursprünglichen Experiment, das Amos und ich durchführten, wählten 73 Prozent der Befragten A bei der Entscheidung 1 und D bei der Entscheidung 2, und nur 3 Prozent präferierten die Kombination von B und C.

Sie wurden gebeten, beide Optionen zu prüfen, ehe Sie Ihre erste Wahl trafen, und das haben Sie vermutlich getan. Aber eines haben Sie bestimmt nicht getan: Sie haben nicht die möglichen Ergebnisse der vier Kombinationen von Wahlen berechnet (A und C, A und D, B und C, B und D), um zu bestimmen, welche Kombination Sie vorziehen. Ihre gesonderten Präferenzen für die beiden Probleme waren intuitiv überzeugend, und es gab keinen Grund, zu erwarten, dass sie zu Problemen führen könnten. Außerdem ist die Verknüpfung der beiden Entscheidungsprobleme eine mühsame Übung, für die Sie Papier und Kuli bräuchten. Sie haben es nicht getan. Betrachten Sie jetzt das folgende Wahlproblem:

AD 25-prozentige Chance, 240 Dollar zu gewinnen, und 75-prozentige Chance, 760 Dollar zu verlieren.

BC 25-prozentige Chance, 250 Dollar zu gewinnen, und 75-prozentige Chance, 750 Dollar zu verlieren.

Diese Wahl ist leicht! Option BC *dominiert* faktisch Option AD (der Fachbegriff für den Sachverhalt, dass eine Option eindeutig besser als die andere ist). Sie wissen bereits, was als Nächstes kommt. Die dominante Option in BC ist die Kombination der beiden verworfenen Optionen im ersten Paar der Entscheidungsprobleme, demjenigen, das nur 3 Prozent der Befragten in unserer ursprünglichen Studie präferierten. Die schwächere Option AD wurde von 73 Prozent der Befragten bevorzugt.[1]

## Weit oder eng?

Diese Serie von Wahlen sagt uns eine Menge über die Grenzen der menschlichen Rationalität. Zum einen hilft sie uns, zu erkennen, was es mit der logischen Konsistenz der Präferenzen von *Humans* auf sich hat – sie ist ein frommer Wunsch. Betrachten Sie noch einmal das letzte Problem, das leichte. Hätten Sie es für möglich gehalten, dass man dieses offensichtliche Wahlproblem in ein Paar von Problemen zerlegen kann, die eine große Mehrzahl von Menschen dazu veranlassen würden, sich für eine schwächere Option zu entscheiden? Dies gilt allgemein: Jede einfache Wahl, die in Kategorien von Gewinnen und Verlusten formuliert ist, lässt sich auf zahllose Weisen in eine Kombination von Wahlen zerlegen, die wahrscheinlich inkonsistente Präferenzen ergeben.

Das Beispiel zeigt auch, dass es kostspielig ist, risikoscheu für Gewinne und risikofreudig für Verluste zu sein. Diese Einstellungen erzeugen die Bereitschaft, eine Prämie zu zahlen, um einen sicheren Gewinn zu erhalten, statt sich auf ein Glücksspiel einzulassen, und auch die Bereitschaft, eine Prämie (auf den Erwartungswert) zu zahlen, um einen sicheren Verlust zu vermeiden. Beide Zahlungen kommen aus derselben Tasche, und wenn man mit beiden Arten von Problemen gleichzeitig konfrontiert ist, sind die diskrepanten Einstellungen vermutlich nicht optimal.

Die Entscheidungen 1 und 2 ließen sich auf zwei verschiedene Weisen konstruieren:

– Ein enges Framing: eine Folge von zwei einfachen Entscheidungen, die getrennt voneinander betrachtet werden
– Ein weites Framing: eine umfassende Entscheidung mit vier Optionen

Eine weite Einrahmung war in diesem Fall offensichtlich überlegen. Tatsächlich ist sie in jedem Fall, in dem mehrere Entscheidungen zusammen in Betracht gezogen werden sollen, überlegen (oder zumindest nicht unterlegen). Stellen Sie sich eine längere Liste von fünf einfachen (binären) Entscheidungen vor, die gleichzeitig erwogen werden sollen. Der weite (umfassende) Rahmen besteht aus einer einzelnen Wahl mit 32 Optionen. Bei einem engen Framing erhält man eine Folge von fünf einfachen Wahlen. Die Folge von fünf Wahlen ist eine der 32 Optionen des

weiten Rahmens. Wird sie die beste sein? Vielleicht, aber nicht sehr wahrscheinlich. Ein rationaler Agent wird sich natürlich für ein weites Framing entscheiden, aber *Humans* neigen von Natur aus zu einer engen Einrahmung.

Wie dieses Beispiel zeigt, kann unser beschränkter Intellekt logische Konsistenz nicht erreichen. Da wir dazu neigen, nur die jeweils aktuell verfügbaren Informationen zu berücksichtigen (WYSIATI-Regel), und mentale Anstrengung scheuen, treffen wir Entscheidungen tendenziell in dem Maße, wie Probleme auftreten, selbst wenn wir eigens aufgefordert werden, sie gemeinsam zu betrachten. Wir haben weder die Neigung noch die mentalen Ressourcen, unsere Präferenzen konsistent zu strukturieren, und unsere Präferenzen sind auch nicht auf magische Weise auf Kohärenz ausgerichtet, wie sie es im Modell des rationalen Agenten sind.

## Samuelsons Problem

Der große Paul Samuelson – ein Gigant unter den Ökonomen des 20. Jahrhunderts – fragte bekanntlich einen Freund, ob er eine Wette auf einen Münzwurf eingehen würde, bei dem er 100 Dollar verlieren oder 200 Dollar gewinnen könnte. Sein Freund antwortete: »Ich werde nicht wetten, weil ich den Verlust von 100 Dollar stärker spüren würde als den Gewinn von 200 Dollar. Aber ich nehme das Angebot an, wenn du versprichst, mich hundert solcher Wetten machen zu lassen.« Sofern Sie kein Entscheidungstheoretiker sind, teilen Sie vermutlich die Intuition von Samuelsons Freund, dass sich das subjektive Risiko verringert, wenn man eine sehr günstige, aber riskante Wette viele Male spielt. Samuelson fand die Antwort seines Freundes interessant und analysierte sie. Er bewies, dass unter einigen sehr spezifischen Bedingungen ein Nutzenmaximierer, der eine einzelne Wette ablehnt, auch das Angebot vieler Wetten ablehnen sollte.

Bemerkenswerterweise schien Samuelson der Tatsache, dass sein Beweis, der selbstverständlich gültig ist, zu einer Schlussfolgerung führte, die dem gesunden Menschenverstand widersprach, keine weitere Beachtung zu schenken: Das Angebot von hundert Wetten ist so verlockend, dass es kein vernünftiger Mensch ablehnen würde. Matthew Rabin und Richard Thaler wiesen darauf hin, dass »eine Lotterie aus hundert Wetten, bei denen man mit gleicher Wahrscheinlichkeit 100 Dollar verlieren oder 200 Dollar gewinnen kann, einen erwarteten Gesamtertrag von 5000 Dollar hat, mit einer Chance von 1 : 2300, überhaupt Geld zu verlieren, und einer Chance von nur 1 : 62 000, mehr als 1000 Dollar zu verlieren«. Sie wollen natürlich darauf hinaus, dass irgendetwas mit der Nutzentheorie als einem Modell rationaler Entscheidung nicht stimmen kann, wenn sie, unter welchen Umständen auch immer, mit solch einer törichten Präferenz vereinbar ist. Samuelson hatte Rabins Beweis der absurden Folgen einer starken Verlustaversion bei kleinen Wetten nicht gekannt, aber er hätte ihn bestimmt nicht überrascht. Seine Bereitschaft, die Möglichkeit überhaupt in Erwägung zu ziehen, dass es rational sein könnte, das »Paket« abzulehnen, belegt die Wirkmächtigkeit des rationalen Modells.

Nehmen wir an, eine sehr einfache Wertfunktion beschreibe die Präferenzen von Samuelsons Freund (nennen wir ihn Sam). Um seine Verlustaversion auszudrücken, schreibt Sam die Wette zuerst um, *nachdem er jeden Verlust mit einem Faktor von zwei multipliziert hat.* Anschließend berechnet er den Erwartungswert der umgeschriebenen Wette. Nachfolgend die Ergebnisse für einen, zwei oder drei Münzwürfe. Sie sind hinreichend aufschlussreich, um eine gewisse pupillenerweiternde Anstrengung zu rechtfertigen.

| | | Erwartungswert |
|---|---|---|
| Ein Wurf | (50 Prozent Verlust 100; 50 Prozent Gewinn 200) | 50 |
| Verluste verdoppelt | (50 Prozent Verlust 200; 50 Prozent Gewinn 200) | 0 |
| Zwei Würfe | (25 Prozent Verlust 200; 50 Prozent Gewinn 100;<br>25 Prozent Gewinn 400) | 100 |
| Verluste verdoppelt | (25 Prozent Verlust 400; 50 Prozent Gewinn 100;<br>25 Prozent Gewinn 400) | 50 |
| Drei Würfe | (12,5 Prozent Verlust 300; 37,5 Prozent Gewinn 0;<br>37,5 Prozent Gewinn 300; 12,5 Prozent Gewinn 600) | 150 |
| Verluste verdoppelt | (12,5 Prozent Verlust 600; 37,5 Prozent Gewinn 0;<br>37,5 Prozent Gewinn 300; 12,5 Prozent Gewinn 600) | 112,5 |

Sie können aus dem Schaubild ersehen, dass der Erwartungswert der Lotterien fünfzig beträgt. Allerdings hat ein Münzwurf für Sam keinen Wert, weil er spürt, dass der Schmerz über den Verlust eines Dollars doppelt so stark ist wie die Freude über den Gewinn eines Dollars. Nachdem Sam die Lotterie so umgeschrieben hat, dass es seine Verlustaversion widerspiegelt, wird er feststellen, dass der Wert der Lotterie null ist.

Betrachten wir jetzt zwei Münzwürfe. Die Chancen, zu verlieren, sind auf 25 Prozent gesunken. Die beiden extremen Ergebnisse (200 verlieren oder 400 gewinnen) heben sich in ihrem Wert gegenseitig auf; sie sind gleich wahrscheinlich, und die Verluste werden doppelt so stark gewichtet wie der Gewinn. Aber das Zwischenergebnis (ein Verlust, ein Gewinn) ist positiv, und das gilt auch für die zusammengesetzte Lotterie insgesamt. Jetzt können Sie die Kosten eines engen Framings und den Zauber der Zusammenführung von Lotterien erkennen. Hier handelt es sich um zwei günstige Lotterien, die einzeln für Sam keinen Wert besitzen. Wenn er dem Angebot bei zwei verschiedenen Gelegenheiten begegnet, wird er es beide Male ablehnen. Doch wenn er die beiden Angebote bündelt, sind sie zusammen 50 Dollar wert!

Noch besser sieht es aus, wenn drei Glücksspiele gebündelt werden. Die extremen Ergebnisse heben sich auch hier weiterhin auf, aber sie sind nicht mehr so bedeutsam. Obwohl der dritte Wurf wertlos ist, wenn er für sich beurteilt wird, hat er den Gesamtwert des Pakets um 62,5 Dollar erhöht. Zu dem Zeitpunkt, da Sam fünf Lotterien angeboten werden, beträgt der Erwartungswert des Angebots 250 Dollar, die Wahrscheinlichkeit, dass er einen Verlust in beliebiger Höhe macht, beträgt 18,75 Prozent, und sein Bargeldäquivalent ist 203,125 Dollar. Der bemerkenswerte Aspekt an dieser Geschichte ist die Tatsache, dass Sam in seiner Verlustaversion nie schwankt. Doch die Zusammenfassung günstiger Lotterien verringert rasch die Verlustwahrscheinlichkeit, und die Auswirkungen der Verlustaversion auf seine Präferenzen verringern sich entsprechend.

Jetzt habe ich einen Sermon für Sam bereit, wenn er das Angebot eines einmaligen außerordentlich günstigen Glücksspiels ablehnt, und für Sie, wenn Sie seine unvernünftige Verlustaversion teilen:

> Ich habe Verständnis für Ihre Abneigung, eine Lotterie zu verlieren, aber das kostet Sie eine Menge Geld.

Betrachten Sie bitte folgende Frage: Liegen Sie auf dem Sterbebett? Ist dies das letzte Angebot einer kleinen, günstigen Lotterie, das Sie jemals in Erwägung ziehen werden?

Natürlich ist es unwahrscheinlich, dass Ihnen genau diese Lotterie ein weiteres Mal angeboten wird, aber Sie werden viele Gelegenheiten haben, attraktive Glücksspiele mit Einsätzen, die im Verhältnis zu Ihrem Vermögen sehr klein sind, in Erwägung zu ziehen. Sie werden sich selbst einen großen finanziellen Gefallen tun, wenn Sie in der Lage sind, jede dieser Lotterien als Teil eines Bündels kleiner Lotterien zu betrachten und das Mantra zu verinnerlichen, das Sie der ökonomischen Rationalität bedeutend näherbringen wird: Sie gewinnen einige, sie verlieren einige. Der Hauptzweck des Mantras ist es, Ihre emotionale Reaktion zu kontrollieren, wenn Sie verlieren. Wenn Sie darauf vertrauen können, dass es effektiv ist, sollten Sie sich daran erinnern, wenn Sie entscheiden, ob Sie ein geringes Risiko mit positivem Erwartungswert eingehen sollen oder nicht. Rufen Sie sich diese Einschränkungen in Erinnerung, wenn Sie das Mantra benutzen:

– Es funktioniert, wenn die Lotterien wirklich unabhängig voneinander sind; es ist nicht anwendbar auf mehrere Investitionen in die gleiche Branche, die alle zusammen in die Hose gehen.
– Es funktioniert nur, wenn der mögliche Verlust nicht so groß ist, dass Sie Angst um Ihr Gesamtvermögen haben müssen. Wenn der Verlust Ihre wirtschaftliche Zukunft erheblich belasten würde, sollten Sie sich vorsehen.
– Es sollte nicht auf hochspekulative Wetten angewendet werden, wo die Gewinnwahrscheinlichkeit für jede Wette sehr niedrig ist.

Wenn Sie die emotionale Disziplin aufbringen, die diese Regel verlangt, werden Sie eine kleine Lotterie niemals isoliert

> betrachten oder für eine kleine Lotterie risikoscheu sein, bis Sie wirklich auf dem Sterbebett liegen – und selbst dann nicht.

Es ist nicht unmöglich, diesen Rat zu beherzigen. Erfahrene Händler auf Finanzmärkten leben Tag für Tag davon, dass sie sich durch weites Framing gegen den Schmerz von Verlusten abschirmen. Wie schon erwähnt, wissen wir jetzt, dass Versuchspersonen (in einem bestimmten Kontext) fast von ihrer Verlustaversion geheilt werden könnten, indem man sie dazu veranlasst, »wie ein Händler zu denken«, so wie erfahrene Baseballkarten-Händler dem Endowment-Effekt nicht so stark unterliegen wie Neulinge. Studenten trafen riskante Entscheidungen (Glücksspiele anzunehmen oder abzulehnen) unter verschiedenen Anweisungen. In der engen Framing-Bedingung wurden sie aufgefordert, »jede Entscheidung so zu treffen, als wäre es die einzige«, und ihre Emotionen zu akzeptieren. Zu den Anweisungen für ein weites Framing einer Entscheidung gehörten Äußerungen wie »Stellen Sie sich vor, Sie wären ein Börsenhändler«, »Sie tun dies ständig« und »Behandeln Sie diese als eine von vielen Anlageentscheidungen, die in ihrer Gesamtheit ein ›Wertpapierdepot‹ ergeben«. Die Experimentatoren beurteilten die emotionalen Reaktionen der Versuchspersonen auf Gewinne und Verluste anhand physiologischer Parameter, wie etwa Veränderungen der elektrischen Leitfähigkeit der Haut, die auch bei der Lügendetektion benutzt werden. Erwartungsgemäß ließ ein weites Framing die emotionale Reaktion auf Verluste abstumpfen und steigerte die Risikobereitschaft.

Die Kombination von Verlustaversion und engem Framing ist ein kostspieliger Fluch. Einzelne Anleger können sich diesem Fluch entziehen, die emotionalen Vorteile eines weiten Framings einstreichen und sich zugleich Zeit und großen Ärger sparen, indem sie die Häufigkeit verringern, mit der sie überprüfen, wie gut sich ihre Anlagen entwickeln. Das zeitnahe Verfolgen täglicher Wertschwankungen ist ein Verlustgeschäft, weil der Ärger über die häufigen kleinen Verluste die Freude über die genauso häufigen kleinen Gewinne überwiegt. Ein Mal im Quartal genügt, und es mag für Privatanleger sogar mehr als genug sein. Das gezielte Ignorieren kurzfristiger Ergebnisse verbessert nicht nur die emotionale Lebensqualität, sondern auch die Qualität der Anlageentscheidungen und -ergebnisse. Die typische kurzfristige Reaktion auf schlechte Neuigkeiten ist eine erhöhte

Verlustaversion. Anleger, die ein zusammenfassendes Feedback erhalten, bekommen solche Nachrichten viel seltener, sind wahrscheinlich weniger risikoscheu und werden langfristig ein höheres Vermögen ansammeln. Sie sind auch weniger anfällig für eine kostspielige ständige Umschichtung ihres Depots, wenn sie nicht wissen, wie sich jede darin vorhandene Aktie tagtäglich (oder wöchentlich oder auch monatlich) entwickelt. Das Bemühen, seine Wertpapierpositionen über mehrere Perioden hinweg nicht zu verändern (was gleichbedeutend ist mit dem »Festlegen« der Anlagegelder), verbessert die Wertentwicklung.[2]

## Risikostrategien

Entscheider, die zu einem engen Framing neigen, konstruieren jedes Mal eine Präferenz, wenn sie vor einer riskanten Wahl stehen. Mit einer Risikostrategie, die sie routinemäßig immer dann anwenden, wenn ein relevantes Problem auftaucht, würden sie besser fahren. Bekannte Beispiele für Risikostrategien sind »Wählen Sie beim Abschluss einer Versicherung immer diejenige mit dem größtmöglichen Selbstbehalt« und »Schließen Sie nie eine erweiterte Produktgarantieversicherung ab«. Eine Risikostrategie ist ein breiter Rahmen. Bei den Versicherungsbeispielen erwarten Sie den gelegentlichen Verlust des gesamten Selbstbehalts oder den gelegentlichen Ausfall eines nicht versicherten Produkts. Entscheidend ist hier Ihre Fähigkeit, den Schmerz über den gelegentlichen Verlust durch den Gedanken zu vermindern oder zu beseitigen, dass die Strategie, die Sie diesem Verlust aussetzte, höchstwahrscheinlich langfristig finanziell vorteilhaft sein wird.

Eine Risikostrategie, die Entscheidungen pauschalisiert, entspricht der Außensicht auf Planungsprobleme, die ich weiter vorn erörtert habe. Die Außenperspektive verschiebt den Fokus von den Besonderheiten der gegenwärtigen Situation auf die statistischen Kenndaten der Ergebnisse in ähnlichen Situationen. Die Außensicht ist ein breiter Rahmen für die Konzipierung von Plänen. Eine Risikostrategie ist ein breiter Rahmen, der eine konkrete risikoträchtige Wahl in eine Gesamtheit ähnlicher Wahlen einbettet.

Die Außensicht und die Risikostrategie sind Abhilfen gegen zwei spezifische Verzerrungen, die viele Entscheidungen betreffen: den übersteigerten Optimismus des Planungsfehlers und die überzogene Vorsicht, die durch die Verlustaversion herbeigeführt wird. Die beiden Verzerrungen wirken antagonistisch. Übersteigerter Optimismus schützt Individuen und Organisationen vor den lähmenden Effekten der Verlustaversion; die Verlustaversion schützt sie vor den Torheiten eines übersteigerten Optimismus. Das Ergebnis ist für den Entscheider eher angenehm. Optimisten glauben, dass ihre Entscheidungen umsichtiger sind, als es tatsächlich der Fall ist, und verlustscheue Entscheider verwerfen zu Recht Optionen mit niedrigem Nutzwert, die sie andernfalls vielleicht

annehmen würden. Es gibt selbstverständlich keine Garantie dafür, dass sich die Verzerrungen in jeder Situation gegenseitig ausgleichen. Eine Organisation, die sowohl übersteigerten Optimismus als auch überzogene Verlustaversion beseitigen kann, sollte dies tun. Die Kombination der Außensicht mit einer Risikostrategie sollte das Ziel sein.

Richard Thaler berichtet von einer Diskussion über Entscheidungsfindung, die er mit den Topmanagern der 25 Sparten eines Konzerns führte. Er bat sie, eine riskante Option zu erwägen, bei der sie mit gleicher Wahrscheinlichkeit einen Großteil des von ihnen kontrollierten Kapitals verlieren oder dieses Kapital verdoppeln könnten. Keiner der Manager war bereit, sich auf ein so riskantes Glücksspiel einzulassen. Anschließend wandte sich Thaler an den Vorstandsvorsitzenden des Konzerns, der ebenfalls anwesend war, und fragte ihn nach seiner Meinung. Ohne zu zögern, antwortete der Vorstandschef: »Ich würde mir wünschen, dass sie alle das Risiko eingehen.« Im Rahmen dieser Diskussion machte sich der Vorstandschef automatisch ein weites Framing zu eigen, das alle 25 Wetten umfasste. Wie Sam, der die Ergebnisse von hundert Münzwürfen sah, konnte er auf die statistische Zusammenfassung setzen, um das Gesamtrisiko abzuschwächen.

## Zum Thema »Risikostrategien«

»Sagen Sie ihr, sie solle wie ein Börsenhändler denken! Bei einigen Geschäften macht man Gewinne, bei einigen macht man Verluste.«

»Ich beschloss, mein Wertpapierdepot nur noch ein Mal pro Quartal zu überprüfen. Ich bin zu verlustscheu, als dass ich angesichts täglicher Kursschwankungen vernünftige Entscheidungen treffen könnte.«

»Sie schließen nie erweiterte Produktgarantieversicherungen ab. Das ist ihre Risikostrategie.«

»Jeder unserer Manager ist in seinem Geschäftsfeld verlustscheu. Das ist verständlich, aber es führt im Endergebnis dazu, dass das Unternehmen zu wenig Risiken eingeht.«

# *32. Buch führen*

Abgesehen von den ganz Armen, für die das Einkommen mit dem Überleben zusammenfällt, sind die Haupttriebkräfte des Strebens nach Geld nicht unbedingt ökonomischer Natur. Für den Milliardär, der noch eine Milliarde reicher werden will, ebenso wie für den Teilnehmer an einem experimentalökonomischen Forschungsprojekt, der ein paar Dollar zusätzlich verdienen will, ist Geld ein Ersatz für Punkte auf einer Skala der Selbstachtung und des Erfolgs. Diese Belohnungen und Bestrafungen, Versprechen und Drohungen sind alle in unserem Kopf. Wir führen sorgfältig Buch darüber. Sie bestimmen unsere Präferenzen und motivieren unsere Handlungen ebenso wie die Anreize, die im gesellschaftlichen Umfeld bereitgestellt werden. Infolgedessen weigern wir uns, Verluste zu begrenzen, wenn dies bedeuten würde, unser Scheitern einzugestehen, schrecken tendenziell vor Handlungen zurück, die wir bereuen könnten, und machen einen trügerischen, aber scharfen Unterschied zwischen Unterlassung und Begehung, Tun und Nichttun, weil das Verantwortungsbewusstsein für das eine größer ist als für das andere. Die »Währung«, die belohnt oder bestraft, ist letztlich oftmals emotionaler Natur, eine Form mentaler Selbstbewertung, die zwangsläufig zu Interessenkonflikten führt, wenn der Einzelne als Stellvertreter für eine Organisation handelt.

## Mentale Buchführung

Seit vielen Jahren faszinieren Richard Thaler Analogien zwischen der Welt der betrieblichen Buchführung und der mentalen Buchführung, die wir dazu benutzen, unser Leben zu organisieren und zu steuern, mit Ergebnissen, die manchmal abstrus und manchmal sehr hilfreich sind. Es gibt mehrere Varianten der mentalen Buchführung. Wir bewahren unser Geld auf verschiedenen Konten auf, die manchmal physisch, manchmal nur mental sind. Wir haben Geld zum Ausgeben, allgemeine Ersparnisse, zweckgebundene Ersparnisse für die Erziehung unserer Kinder oder für medizinische Notfälle. Es gibt eine klare Hierarchie in unserer Bereitschaft, diese Konten in Anspruch zu nehmen, um laufende Bedürfnisse zu befriedigen. Wir benutzen Konten zu Zwecken der Selbstkontrolle, etwa wenn wir ein Haushaltsbudget aufstellen, den täglichen Verbrauch von Espressos begrenzen oder uns mehr Zeit für sportliche Aktivitäten nehmen. Oftmals bezahlen wir für Selbstkontrolle, zum Beispiel indem wir gleichzeitig Geld auf ein Sparkonto einzahlen und unsere Kreditkarten belasten. Die *Econs* im Modell des rationalen Agenten greifen nicht auf mentale Buchführung zurück; sie haben einen umfassenden Überblick über die Ergebnisse und werden von externen Anreizen angetrieben. Für die *Humans* sind mentale Bücher eine Form des engen Framings; sie sorgen dafür, dass alles von einem begrenzten Intellekt beherrschbar bleibt.

Mentale Buchführung ist weit verbreitet. Erinnern wir uns daran, dass Profigolfer erfolgreicher putten, wenn sie versuchen, einen Bogey zu vermeiden, statt einen Birdie zu erzielen. Eine Schlussfolgerung, die wir daraus ziehen können, ist, dass die besten Golfer für jedes Loch getrennt Buch führen; sie führen nicht nur Buch über ihren Gesamterfolg. Ein ironisches Beispiel, über das Thaler in einem frühen Artikel schreibt, veranschaulicht besonders gut, wie sich die mentale Buchführung auf das Verhalten auswirkt:

> Zwei begeisterte Sportfans wollen 40 Meilen fahren, um sich ein Basketballspiel anzusehen. Einer von ihnen hat seine Eintrittskarte bezahlt; der andere war gerade unterwegs, um

> sich eine Eintrittskarte zu kaufen, als er von einem Freund eine geschenkt bekam. Für den Abend, an dem das Spiel stattfinden soll, ist ein Schneesturm angesagt. Welcher der beiden Ticketbesitzer wird dem Schneesturm eher trotzen?

Die Antwort kommt sofort: Wir wissen, dass der Fan, der für seine Eintrittskarte bezahlt hat, eher fahren wird. Mentale Buchführung liefert die Erklärung. Wir nehmen an, dass beide Fans ein Konto für das Spiel eröffneten, das sie sehen wollten. Das Versäumen des Spiels schließt die Konten mit einem negativen Saldo. Unabhängig davon, wie sie an ihre Eintrittskarte kamen, werden beide enttäuscht sein – aber der Endsaldo ist für denjenigen, der eine Eintrittskarte kaufte und jetzt blank ist und um das Spiel gebracht wurde, deutlich negativer. Weil es für diese Person schlimmer ist, zu Hause zu bleiben, ist er stärker motiviert, sich das Spiel anzusehen, und wird daher eher den Versuch unternehmen, duch einen Schneesturm durch zu fahren.[1] Dies sind stillschweigende Berechnungen des emotionalen Saldos, wie sie System 1 ohne gezielte Überlegung ausführt. Die Emotionen, die Menschen mit dem Zustand ihrer mentalen Konten verknüpfen, werden in der ökonomischen Standardtheorie nicht berücksichtigt. Ein *Econ* würde erkennen, dass die Eintrittskarte bereits bezahlt wurde und nicht zurückgegeben werden kann. Ihre Kosten sind »versunken«, und dem *Econ* wäre es egal, ob er die Eintrittskarte für das Spiel gekauft hat oder sie von einem Freund bekam (wenn *Econs* Freunde haben). Um dieses rationale Verhalten umzusetzen, müsste sich System 2 der kontrafaktischen Möglichkeit bewusst sein: »Würde ich auch dann noch in den Schneesturm fahren, wenn ich die Eintrittskarte gratis von einem Freund bekommen hätte?« Es bedarf eines aktiven und disziplinierten Intellekts, um eine so schwierige Frage aufzuwerfen. Zu einem ähnlichen Fehler neigen Privatanleger, wenn sie Aktien aus ihrem Wertpapierdepot verkaufen:

> Sie brauchen Geld, um die Kosten für die Hochzeit Ihrer Tochter zu bestreiten, und werden deshalb einige Aktien verkaufen. Sie erinnern sich, zu welchem Kurs Sie jede Aktie

> gekauft haben, und können sie als einen »Gewinner«, der gegenwärtig mehr wert ist, als Sie dafür bezahlt haben, oder als einen »Verlierer« identifizieren. Unter den Aktien, die Ihnen gehören, ist Blueberry Tiles ein Gewinner; wenn Sie Ihren Bestand heute verkaufen, machen Sie einen Gewinn von 5000 Dollar. Außerdem halten Sie Aktien von Tiffany Motors, die gegenwärtig weniger wert sind, als Sie dafür bezahlt haben. Der Kurs beider Aktien ist in den letzten beiden Wochen stabil geblieben. Welche werden Sie eher verkaufen?

Eine plausible Formulierung dieses Entscheidungsproblems lautet folgendermaßen: »Ich könnte das Blueberry-Tiles-Konto schließen und einen Erfolg in meiner Bilanz als Anleger verbuchen. Wahlweise könnte ich auch das Tiffany-Motors-Konto schließen und einen Misserfolg in meiner Anlegerbilanz verbuchen. Was sollte ich eher tun?« Wenn das Problem als eine Wahl zwischen Freude und Ärger formuliert wird, werden Sie sicherlich die Blueberry Tiles verkaufen und sich über Ihre hervorragenden Investment-Fähigkeiten freuen. Wie zu erwarten, hat die finanzwissenschaftliche Forschung eine massive Präferenz für den Verkauf von Gewinnern statt von Verlierern dokumentiert – eine Verzerrung, die eine nicht unmittelbar verständliche Bezeichnung erhalten hat: der Dispositionseffekt.[2]

Der Dispositionseffekt ist ein Fall von engem Framing. Der Anleger hat ein Konto für jede Aktie, die er gekauft hat, eingerichtet, und er will jedes Konto mit einem Gewinn abschließen. Ein rationaler Agent hätte einen umfassenden Überblick über den Wertpapierbestand und würde die Aktie verkaufen, die in Zukunft mit der geringsten Wahrscheinlichkeit eine gute Wertentwicklung verzeichnen wird, ohne zu überlegen, ob es sich um einen Gewinner oder einen Verlierer handelt. Amos erzählte mir von einem Gespräch mit einem Finanzberater, der ihn um eine vollständige Liste der Aktien in seinem Depot einschließlich ihrer jeweiligen Kaufkurse bat. Der Berater zog ein erstauntes Gesicht, als Amos vorsichtig fragte: »Sollte das nicht egal sein?« Offensichtlich war er von jeher der Auffassung gewesen, dass der Zustand des mentalen Kontos ein relevanter Gesichtspunkt sei.

Amos’ Vermutung über die Ansichten des Finanzberaters war vermutlich richtig, aber er hatte unrecht, den Kaufpreis als irrelevant abzutun. Der

Kaufpreis spielt eine Rolle und sollte auch von *Econs* berücksichtigt werden. Der Dispositionseffekt ist eine kostspielige Verzerrung, weil die Frage, ob man Gewinner oder Verlierer verkaufen soll, ein eindeutige Antwort hat, und es ist nicht so, dass es egal wäre. Wenn Ihnen mehr an Ihrem Vermögen als an Ihren unmittelbaren Gefühlen gelegen ist, sollten Sie den Verlierer Tiffany Motors verkaufen und an den gewinnenden Blueberry Tiles festhalten. Zumindest in den Vereinigten Staaten liefern Steuern einen starken Anreiz: Die Realisierung von Verlusten vermindert Ihre Steuerlast, während Ihnen die Veräußerung von Gewinnern zusätzliche Steuern aufbürdet. Diese grundlegende Tatsache über Finanztransaktionen ist allen amerikanischen Anlegern bekannt, und sie bestimmt die Entscheidungen, die sie während eines Monats im Jahr treffen – Anleger stoßen im Dezember, wenn sie an ihre Steuern denken, mehr Verlierer ab. Der Steuervorteil ist natürlich das ganze Jahr über verfügbar, aber die übrigen elf Monate des Jahres setzt sich die mentale Buchhaltung gegenüber dem gesunden Menschenverstand in finanziellen Angelegenheiten durch. Ein weiteres Argument gegen den Verkauf von Gewinnern ist die gut dokumentierte Marktanomalie, wonach Aktien, deren Kurs in jüngster Vergangenheit gestiegen ist, wahrscheinlich zumindest noch eine kurze Weile weitere Kursgewinne verzeichnen werden. Der Nettoeffekt ist groß: Der erwartete Mehrertrag nach Steuern des Verkaufs von Tiffany-Aktien im Vergleich zu Blueberry-Aktien beläuft sich auf 3,4 Prozent über das nächste Jahr. Ein mentales Konto mit einem Gewinn abzuschließen, erzeugt ein Lustgefühl, aber es ist ein Lustgefühl, für das man bezahlt. Ein *Econ* würde diesen Fehler nie machen, und erfahrene Anleger, die ihr System 2 benutzen, sind dafür weniger anfällig als Neulinge.[3]

Ein rationaler Entscheider interessiert sich nur für die zukünftigen Konsequenzen gegenwärtiger Investments. Zurückliegende Fehler zu rechtfertigen ist für den *Econ* nicht von Belang. Die Entscheidung, zusätzliche Finanzmittel in ein Verlustgeschäft zu investieren, wenn bessere Anlagemöglichkeiten zur Verfügung stehen, wird »Fehlschluss aus versunkenen, irreversiblen Kosten« *(sunkcost fallacy)* genannt, ein kostspieliger Fehler, der bei großen und kleinen Entscheidungen beobachtet wird. In einen Schneesturm hineinzufahren, weil man Eintrittskarten bezahlt hat, ist ein Fehler aus versunkenen Kosten.

Stellen Sie sich ein Unternehmen vor, das bereits 50 Millionen Dollar für ein Projekt ausgegeben hat. Das Projekt ist jetzt im Rückstand, und die Prognosen über seine letztendliche Rentabilität sind ungünstiger als im anfänglichen Planungsstadium. Es bedarf einer zusätzlichen Investition von 60 Millionen Dollar, um dem Projekt eine Chance zu geben. Alternativ könnte die gleiche Summe auch in ein neues Projekt investiert werden, das nach jetzigem Stand vermutlich rentabler sein wird. Was wird das Unternehmen tun? Allzu oft fährt ein Unternehmen, das von versunkenen Kosten betroffen ist, in den Schneesturm hinein und wirft gutes Geld schlechtem Geld hinterher, statt sich mit der Demütigung abzufinden, das Konto eines kostspieligen Fehlschlags zu schließen. Diese Situation ist im rechten oberen Feld des viergeteilten Musters dargestellt (Seite 389), wo es darum geht, zwischen einem sicheren Verlust und einer ungünstigen Lotterie zu wählen, die unklugerweise oftmals bevorzugt wird.

Die »Steigerung des Einsatzes« *(escalation of commitment)* bei zum Scheitern verurteilten Projekten ist ein Fehler aus der Perspektive der Firma, aber nicht unbedingt aus der Sicht des Managers, der für ein fehlgeschlagenes Projekt zuständig ist und sich damit identifiziert hat. Das Aufgeben des Projekts wird einen dauerhaften Makel in der Erfolgsbilanz des Managers erzeugen, und seinen persönlichen Interessen ist vielleicht am besten gedient, wenn er weiter die Finanzmittel der Organisation aufs Spiel setzt, in der Hoffnung, die ursprüngliche Investition wieder hereinzuholen – oder zumindest in dem Bestreben, den Tag der Abrechnung hinauszuschieben. Wenn versunkene Kosten entstanden sind, decken sich die Anreize des Managers nicht mit den Zielen des Unternehmens und seiner Anteilseigner – dies ist ein gängiges Beispiel für das sogenannte »Stellvertreter-Problem« *(agency problem).* Aufsichtsräte sind sich dieser Konflikte durchaus bewusst und lösen oftmals einen Vorstandschef ab, der durch frühere Entscheidungen belastet ist und nur widerwillig Verluste zurückführt. Die Mitglieder des Aufsichtsrats sind nicht unbedingt der Ansicht, der neue Vorstandsvorsitzende wäre kompetenter als sein Vorgänger. Aber Sie wissen, dass er nicht dieselben mentalen Konten führt und daher besser in der Lage ist, bei der Beurteilung aktueller Geschäftschancen die versunkenen Kosten vergangener Entscheidungen zu ignorieren.

Der Fehlschluss aus versunkenen Kosten sorgt dafür, dass Menschen zu lange an ungeliebten Arbeitsplätzen ausharren, in unglücklichen Ehen

bleiben und an aussichtslosen Forschungsprojekten festhalten. Ich habe oft junge Wissenschaftler dabei beobachtet, wie sie mit aller Gewalt ein zum Scheitern verurteiltes Projekt zu retten versuchten, wenn sie besser beraten gewesen wären, es aufzugeben und mit einem neuen zu beginnen. Die Forschungen deuten glücklicherweise darauf hin, dass der Fehlschluss zumindest in einigen Kontexten überwunden werden kann.[4] Sowohl in volkswirtschaftlichen als auch in betriebswirtschaftlichen Seminaren wird der Fehlschluss aus versunkenen Kosten als ein Fehler dargestellt und gelehrt, ganz offenbar mit einer positiven Wirkung: Es gibt Hinweise darauf, dass Absolventen dieser Fächer eher als andere bereit sind, ein scheiterndes Projekt aufzugeben.

## Reue

Reue ist eine Emotion, und sie ist ebenso eine Bestrafung, die wir uns selbst zufügen. Die Furcht davor, zu bereuen, ist ein Faktor bei vielen der Entscheidungen, die Menschen treffen (»Lass das, du wirst es sonst bereuen« ist eine weitverbreitete Warnung), und den meisten Menschen ist dieses Gefühl vertraut. Zwei niederländische Psychologen haben diesen emotionalen Zustand gut beschrieben; sie wiesen darauf hin, dass Reue »mit dem Gefühl einhergeht, dass man es hätte besser wissen müssen, mit einem Gefühl der Bangigkeit, mit Gedanken an die Fehler, die man begangen hat, und die Chancen, die man ungenutzt verstreichen ließ, mit einer Tendenz, sich selbst einen Tritt zu geben und seinen Fehler zu korrigieren, und mit dem Wunsch, das Ereignis ungeschehen zu machen und eine zweite Chance zu bekommen«.[5] Intensives Bedauern erlebt man, wenn man sich vorstellen kann, dass man sich leicht anders hätte verhalten können.

Reue ist eine der kontrafaktischen Emotionen, die durch die Verfügbarkeit von Alternativen zur Wirklichkeit ausgelöst werden. Nach jedem Flugzeugabsturz gibt es Geschichten über Passagiere, die nicht an Bord hätten »sein sollen« – sie bekamen im letzten Moment einen Sitz, sie wurden von einer anderen Fluggesellschaft umgebucht oder sie sollten einen Tag früher fliegen, mussten den Flug aber verschieben. Das gemeinsame Merkmal dieser erschütternden Geschichten ist die Tatsache, dass sie sich um ungewöhnliche Ereignisse drehen – und ungewöhnliche Ereignisse sind in der Fantasie leichter ungeschehen zu machen als normale Ereignisse. Das assoziative Gedächtnis enthält eine Repräsentation der normalen Welt und ihrer Regeln. Ein ungewöhnliches Ereignis zieht die Aufmerksamkeit auf sich, und es aktiviert auch die Vorstellung von dem Ereignis, das unter denselben Umständen normal gewesen wäre.

Um die Verbindung zwischen Reue und Normalität zu verstehen, betrachten Sie das folgende Szenario:[6]

Mr Brown nimmt fast nie Anhalter mit. Gestern nahm er einen Mann mit und wurde ausgeraubt.

Mr Smith nimmt häufig Anhalter mit. Gestern nahm er einen Mann mit und wurde ausgeraubt.

Wer von den beiden wird den Vorfall stärker bereuen?

Die Ergebnisse sind nicht überraschend: 88 Prozent der Befragten sagten, Mr Brown, 12 Prozent nannten Mr Smith.

Reue ist nicht das Gleiche wie Tadel. Anderen Teilnehmern wurde diese Frage über denselben Zwischenfall gestellt:

Wer wird schärfer von anderen kritisiert werden?

Die Ergebnisse: Mr Brown 23 Prozent, Mr Smith 77 Prozent. Reue und Tadel werden beide durch den Vergleich mit einer Norm hervorgerufen, aber die relevanten Normen unterscheiden sich. Die Emotionen, die Mr Brown und Mr Smith erleben, werden von dem bestimmt, was sie für gewöhnlich mit Anhaltern tun. Das Mitnehmen eines Anhalters ist ein ungewöhnliches Ereignis für Mr Brown, daher erwarten die meisten Menschen, dass er ein stärkeres Gefühl der Reue empfindet. Ein urteilender Beobachter dagegen wird beide Männer an herkömmlichen Normen vernünftigen Verhaltens messen und wahrscheinlich Mr Smith dafür tadeln, dass er gewohnheitsmäßig unvertretbar hohe Risiken eingeht.[7] Wir sind versucht zu sagen, dass Mr Smith sein Schicksal verdiente und dass Mr Brown Pech hatte. Aber Mr Brown ist derjenige, der sich mit größerer

Wahrscheinlichkeit über sich selbst ärgern wird, weil er sich in diesem einen Fall untypisch verhielt.

Entscheider wissen, dass sie zu Reue neigen, und das Vorwegnehmen dieser schmerzlichen Emotion spielt bei vielen Entscheidungen eine Rolle. Intuitionen über Reue sind bemerkenswert gleichförmig und zwingend, wie das nächste Beispiel verdeutlicht.[8]

> Paul besitzt Aktien des Unternehmens A. Im letzten Jahr überlegte er, ob er auf Aktien des Unternehmens B umsteigen sollte, aber er entschied sich dagegen. Jetzt erfährt er, dass er um 1200 Dollar reicher wäre, wenn er auf Aktien des Unternehmens B umgestiegen wäre.
>
> George besaß Aktien des Unternehmens B. Im vergangenen Jahr tauschte er diese gegen Aktien des Unternehmens A aus. Er erfährt jetzt, dass er um 1200 Dollar reicher wäre, wenn er seine Aktien an Unternehmen B behalten hätte.
>
> Wer bereut sein Verhalten stärker?

Die Ergebnisse sind eindeutig: 8 Prozent der Befragten sagen, Paul, 92 Prozent sagen, George. Dies ist seltsam, weil die Situationen der beiden Anleger objektiv gesehen gleich sind. Beide besitzen A-Aktien, und beide wären um die gleiche Summe reicher, wenn sie B-Aktien besitzen würden. Der einzige Unterschied besteht darin, dass George durch Handeln in diese Lage kam, während Paul durch Nichthandeln in die gleiche Lage kam. Dieses kurze Beispiel verdeutlicht ein allgemeines Muster: Menschen erwarten, stärkere emotionale Reaktionen (einschließlich Reue) auf ein Ergebnis zu zeigen, das durch Tun zustande kommt, als auf das gleiche Ergebnis, wenn es durch Untätigkeit zustande kommt. Dies wurde bei Glücksspielen bestätigt: Menschen erwarten, zufriedener zu sein, wenn sie

riskant spielen und gewinnen, als wenn sie auf Glücksspiele verzichten und die gleiche Summe so erhalten. Die Asymmetrie ist für Verluste mindestens genauso stark, und sie gilt für Tadel wie für Reue.[9] Das Entscheidende ist nicht der Unterschied zwischen Begehung und Unterlassung, sondern die Unterscheidung zwischen Standardoptionen und Handlungen, die vom Standard abweichen.[10] Wenn man vom Standard abweicht, kann man sich leicht die Norm vorstellen – und wenn die Standardoption mit negativen Konsequenzen assoziiert ist, kann die Diskrepanz zwischen den beiden die Ursache schmerzlicher Gefühle sein. Die Standardoption für den Aktienbesitzer besteht darin, die Aktie nicht zu verkaufen, aber wenn Sie Ihrem Kollegen am Morgen über den Weg laufen, besteht die Standardoption darin, ihn zu grüßen. Eine Aktie zu verkaufen und einen Kollegen nicht zu grüßen sind beides Abweichungen von der Standardoption und natürliche Kandidaten für Reue und Tadel.

Teilnehmer einer Studie haben die Wirkmächtigkeit von Standardoptionen in bezwingender Weise verdeutlicht, als sie eine Computersimulation von Blackjack spielten. Einige Spieler wurden gefragt: »Wünschen Sie eine weitere Karte?« Während andere gefragt wurden: »Wünschen Sie keine weitere Karte?« Unabhängig von der Frage war das Bejahen mit einer viel intensiveren Reue verbunden als das Verneinen, wenn das Ergebnis negativ war! Die Frage legt offensichtlich eine Standardantwort nahe, die lautet: »Ich habe keinen starken Wunsch, dies zu tun.« Die Abweichung vom Standard erzeugt Reue. Eine andere Situation, in der Tun der Standard ist, ist die eines Trainers, dessen Mannschaft ihr letztes Spiel haushoch verloren hat. Vom Trainer wird erwartet, Spieler auszuwechseln oder die Strategie zu ändern, und wenn er das nicht tut, wird dies zu Tadel und Reue führen.[11]

Die Asymmetrie im Risiko des Bereuens begünstigt konventionelle und risikoscheue Wahlen. Die Verzerrung tritt in vielen Kontexten in Erscheinung. Verbraucher, die daran erinnert werden, dass sie ihre Kaufentscheidungen möglicherweise bereuen werden, zeigen eine erhöhte Präferenz für konventionelle Optionen und ziehen Markenartikel No-Name-Produkten vor.[12] Das Verhalten der Manager von Investmentfonds gegen Ende des Geschäftsjahres zeigt einen Effekt vorweggenommener Bewertung: Sie neigen dazu, ihren Wertpapierbestand um ungewöhnliche

und anderweitig fragwürdige Aktien zu bereinigen.[13] Selbst Entscheidungen über Leben und Tod können davon beeinflusst sein. Stellen Sie sich einen Arzt mit einem schwer kranken Patienten vor. Die eine Behandlung entspricht den therapeutischen Richtlinien, die andere ist unkonventionell. Der Arzt hat gewisse Anhaltspunkte dafür, dass die unkonventionelle Behandlung die Überlebenschancen des Patienten verbessern könnte, aber die empirische Evidenz ist nicht eindeutig. Der Arzt, der die unkonventionelle Behandlung verschreibt, sieht sich einem erheblichen Risiko von Reue und Tadel ausgesetzt und riskiert vielleicht sogar einen Prozess. In der Rückschau wird sich die normale Wahl leichter vorstellen lassen; die ungewöhnliche Wahl wird schnell vergessen. Es stimmt zwar, dass ein gutes Ergebnis das Renommee des wagemutigen Arztes verbessern wird, aber der potenzielle Nutzen ist kleiner als die potenziellen Kosten, weil der Erfolg im Allgemeinen ein »normaleres« Ergebnis ist als der Misserfolg.

## Verantwortung

In etlichen Kontexten werden Verluste etwa doppelt so stark gewichtet wie Gewinne: bei der Wahl zwischen Glücksspielen, beim Endowment-Effekt und bei Reaktionen auf Preisänderungen. Der Verlustaversionskoeffizient ist in einigen Situationen noch viel höher. So mögen wir für Aspekte unseres Lebens, die wichtiger als Geld sind, etwa Gesundheit, eine stärkere Verlustaversion besitzen. [14] Außerdem steigt unser Widerwille, wichtige Besitztümer zu »verkaufen«, drastisch an, wenn wir dadurch für ein schreckliches Ergebnis verantwortlich werden. Richard Thalers früher Klassiker über das Verbraucherverhalten enthielt ein sehr anschauliches Beispiel, das in der folgenden Frage leicht abgewandelt ist:[15]

> Sie sind mit einer Infektionsquelle in Kontakt gekommen, und die Krankheit, die Sie sich dabei zugezogen haben könnten, führt innerhalb einer Woche zu einem schnellen, schmerzlosen Tod. Die Wahrscheinlichkeit, dass Sie sich infiziert haben, beträgt 1:1000. Es gibt einen Impfstoff, der nur so lange wirksam ist, wie noch keine Symptome aufgetreten sind. Was wäre der Höchstpreis, den Sie für den Impfstoff zu zahlen bereit wären?

Die meisten Menschen sind bereit, einen erheblichen, aber begrenzten Betrag zu zahlen. Der Möglichkeit des Todes ins Auge zu sehen, ist unangenehm, aber das Risiko ist gering, und es erscheint unvernünftig, sich selbst zu ruinieren, um es zu vermeiden. Betrachten wir nun eine kleine Variation:

> Für ein Forschungsprojekt über die oben erwähnte Erkrankung werden Freiwillige gesucht. Sie müssen nicht mehr tun, als sich einem Ansteckungsrisiko von 1 : 1000 auszusetzen. Was ist die

> Mindestsumme, die Sie verlangen würden, um freiwillig an diesem Programm teilzunehmen? (Es würde Ihnen nicht erlaubt sein, den Impfstoff zu kaufen.)

Erwartungsgemäß ist die Vergütung, die Freiwillige für die Teilnahme fordern, viel höher als der Preis, den sie für den Impfstoff zu zahlen bereit wären. Thaler berichtete, dass ein typisches Verhältnis etwa 50:1 beträgt. In dem extrem hohen Verkaufspreis spiegeln sich zwei Merkmale dieses Problems wider. Erstens sollen Sie Ihre Gesundheit nicht verkaufen; die Transaktion gilt als sittenwidrig, und der Widerwille, sich darauf einzulassen, schlägt sich in einem höheren Preis nieder. Vielleicht am wichtigsten aber ist, dass Sie für das Ergebnis, wenn es negativ sein sollte, verantwortlich sein werden. Sie wissen, dass Sie, wenn Sie eines Morgens mit Symptomen aufwachen werden, die Ihren baldigen Tod ankündigen, im zweiten Fall mehr Reue empfinden werden als im ersten. Schließlich hätten Sie das Angebot, Ihre Gesundheit zu verkaufen, ablehnen können, ohne auch nur einen Gedanken an den Preis zu verschwenden. Sie hätten bei der Standardoption bleiben und nichts tun können, und jetzt wird Sie dieser kontrafaktische Gedanke für den Rest Ihres Lebens verfolgen.

Die früher erwähnte Erhebung über die Reaktionen von Eltern auf ein potenziell gefährliches Insektizid umfasste auch eine Frage nach der Bereitschaft, ein erhöhtes Risiko einzugehen. Die Befragten sollten sich vorstellen, sie benutzten ein Insektizid, bei dem das Risiko einer Inhalations- und Kleinkindvergiftung 15 pro 10 000 Dosen betrug. Ein weniger teures Insektizid, für das das Risiko von 15 auf 16 von 10 000 Dosen stieg, stand ebenfalls zur Verfügung. Die Eltern wurden nach dem Preisnachlass gefragt, der sie dazu veranlassen würde, auf das weniger kostspielige (und weniger sichere) Produkt umzusteigen. Mehr als zwei Drittel der Eltern in der Studie antworteten, sie würden keinesfalls das neue Produkt kaufen, egal zu welchem Preis! Schon die bloße Vorstellung, die Gesundheit ihres Kindes gegen Geld einzutauschen, empörte sie. Die Minderheit, die einen für sich akzeptablen Preisabschlag fand, verlangte einen Betrag, der erheblich höher war als die Summe, die sie für eine weit stärkere Verbesserung der gesundheitlichen Sicherheit des Produktes zu zahlen bereit war.

Jeder kann das Widerstreben von Eltern verstehen und nachempfinden, selbst ein nur geringfügig erhöhtes gesundheitliches Risiko für ihr Kind gegen Geld einzutauschen. Man sollte jedoch bedenken, dass diese Einstellung inkohärent und potenziell schädlich für das gesundheitliche Wohl derjenigen ist, die wir beschützen wollen. Selbst die liebevollsten Eltern haben begrenzte Ressourcen an Zeit und Geld, um ihr Kind zu beschützen (das mentale Konto »die Sicherheit meines Kindes schützen« hat ein begrenztes Budget), und es erscheint vernünftig, diese Ressourcen in einer Weise zu verwenden, die den größtmöglichen Nutzen bringt. Geld, das man dadurch sparen könnte, dass man eine geringfügige Erhöhung des gesundheitlichen Risikos durch ein Pestizid in Kauf nimmt, ließe sich zweifellos nutzbringender ausgeben, um andere Schadensrisiken für das Kind zu verringern, etwa durch den Kauf eines sichereren Kindersitzes oder von Kindersicherungen für Steckdosen. Der *taboo trade-off* – eine »tabuisierte Kosten-Nutzen-Abwägung« – gegen jegliche Risikoerhöhung stellt keine effiziente Nutzung des Sicherheitsbudgets dar.[16] Tatsächlich mag der Widerstand stärker durch eine egoistische Furcht vor Reue als durch den Wunsch motiviert sein, die Sicherheit des Kindes zu optimieren. Der »Was wenn?«-Gedanke, der sich allen Eltern aufdrängt, die einen solchen »Handel« eingehen, ist ein Bild der Reue und Scham, die sie empfinden würden, falls das Pestizid einen Gesundheitsschaden verursachen sollte.

Die intensive Abneigung dagegen, ein erhöhtes Risiko gegen einen anderen Vorteil einzutauschen, manifestiert sich im großen Maßstab in den gesetzlichen Regelungen zur Risikobeschränkung. Dieser Trend ist besonders stark in Europa, wo das Vorsichtsprinzip, das jegliche Handlung verbietet, die andere schädigen könnte, ein allgemein anerkannter Grundsatz ist.[17] Im Kontext der Rechtsordnung verlangt das Vorsichtsprinzip von jedem, der Handlungen unternimmt, die Menschen oder der Umwelt schaden könnten, den zweifelsfreien Nachweis der Unbedenklichkeit. Mehrere internationale Organisationen haben dargelegt, dass der fehlende wissenschaftliche Nachweis möglicher Schadensrisiken keine hinreichende Rechtfertigung für das Eingehen von Risiken ist. Der Jurist Cass Sunstein weist darauf hin, dass das Vorsichtsprinzip kostspielig ist und bei enger Auslegung den Fortschritt lähmen kann. Er erwähnt eine eindrucksvolle Reihe von Innovationen, die den Test nicht bestanden hätten,

darunter »Flugzeuge, Klimaanlagen, Antibiotika, Autos, Chlor, der Impfstoff gegen Masern, Operationen am offenen Herzen, Radio, Kältetechnik, der Windpocken-Impfstoff und Röntgenstrahlen«. Die starke Version des Vorsichtsprinzips ist offensichtlich unhaltbar. Aber eine verstärkte Verlustaversion ist in eine starke und weitverbreitete moralische Intuition integriert; sie hat ihren Ursprung in System 1. Das Dilemma zwischen stark verlustaversen moralischen Einstellungen und effizientem Risikomanagement hat keine einfache und zwingende Lösung.

Wir verbringen einen Großteil des Tages damit, das emotionale Unbehagen, das wir uns selbst zufügen, vorherzusehen und möglichst zu vermeiden. Wie ernst sollten wir diese immateriellen Ergebnisse nehmen, die Selbstbestrafungen (und gelegentlichen -belohnungen), die wir erleben, wenn wir über unser Leben Buch führen? Die *Econs* sollten sie nicht haben, und für die *Humans* sind sie kostspielig. Sie führen zu Handlungen, die dem Vermögen von Individuen, der Solidität der Politik und dem Wohl der Gesellschaft schaden. Aber die Emotionen der Reue und moralischen Verantwortung sind real, und die Tatsache, dass *Econs* sie nicht haben, ist möglicherweise nicht von Belang.

Ist es insbesondere vernünftig, seine Entscheidungen von der Antizipation zukünftiger Reue beeinflussen zu lassen? Die Empfänglichkeit für Reuegefühle ist wie die Anfälligkeit für Ohnmachtsanfälle eine Tatsache des Lebens, an die man sich anpassen muss. Ein hinreichend vermögender, vorsichtiger Anleger mag sich den Luxus eines Wertpapierdepots leisten können, das die Reue-Erwartung minimiert, auch wenn es den Vermögenszuwachs nicht maximiert.

Man kann auch Vorsichtsmaßnahmen ergreifen, die einen vor Reue schützen. Am nützlichsten ist es vielleicht, ausdrücklich auf die Erwartung der Reue einzugehen. Wenn etwas schiefläuft und Sie sich daran erinnern können, dass Sie die Möglichkeit der Reue sorgfältig erwogen haben, bevor Sie eine Entscheidung trafen, werden Sie wahrscheinlich weniger Reue erleben. Sie sollten auch wissen, dass Reue und der Rückschaufehler zusammen auftreten, sodass vermutlich alles hilfreich ist, was Sie tun können, um sich von der Rückschau abzuhalten. Meine persönliche Rückschau-Vermeidungsstrategie besteht darin, bei einer Entscheidung mit langfristigen Folgen entweder sehr gründlich oder völlig unbekümmert zu sein. Die Rückschau ist schlimmer, wenn man ein bisschen nachdenkt,

gerade so viel, dass man sich später sagen muss: »Ich hätte beinahe eine bessere Entscheidung getroffen.«

Daniel Gilbert und seine Kollegen stellten die provokante Behauptung auf, dass Menschen im Allgemeinen mehr Reue erwarten, als sie tatsächlich erleben werden, weil sie die Wirksamkeit der psychologischen Abwehrmechanismen, die sie einsetzen, unterschätzen. Die beiden Wissenschaftler nennen dies das »psychologische Immunsystem«.[18] Sie empfehlen, der Reue nicht allzu viel Gewicht beizumessen; selbst wenn man bereuen sollte, wird es weniger wehtun, als man heute glaubt.

## Zum Thema »Mentale Buchführung«

»Er hat getrennte mentale Konten für Bar- und Kreditkäufe. Ich erinnere ihn immer wieder daran, dass Geld Geld ist.«

»Wir halten nur deshalb an dieser Aktie fest, weil wir unser mentales Konto nicht mit einem Verlust abschließen wollen. Das ist der Dispositionseffekt.«

»Wir haben ein hervorragendes Gericht in einem Restaurant entdeckt, und wir probieren keine andere Speise, um uns das Gefühl der Reue zu ersparen.«

»Der Verkäufer zeigte mir den teuersten Kindersitz und sagte, es sei der sicherste, und ich konnte mich nicht dazu durchringen, das billigere Modell zu kaufen. Es fühlte sich an wie ein *taboo trade-off*.«

# *33. Umkehrungen*

Sie sollen die Höhe des Schadensersatzes für die Opfer von Gewaltverbrechen festsetzen. Sie prüfen den Fall eines Mannes, der aufgrund seiner Schusswunde seinen rechten Arm nicht mehr gebrauchen kann. Er wurde angeschossen, als er in einen Raubüberfall geriet, der sich in einem Nachbarschaftsladen in der Nähe ereignete.

Zwei Läden befanden sich in der Nähe der Wohnung des Opfers, wobei der Mann einen davon regelmäßiger aufsuchte als den anderen. Betrachten wir zwei Szenarien:

a. Der Raubüberfall ereignete sich in dem Geschäft, das der Mann regelmäßig aufsuchte.
b. Das Stammgeschäft des Mannes war wegen einer Beisetzung geschlossen, sodass er in dem anderen Laden einkaufte, wo er angeschossen wurde.

Sollte es sich auf die Höhe des Schadensersatzes auswirken, in welchem Geschäft der Mann angeschossen wurde?[1]

Sie treffen Ihr Urteil in einer gemeinsamen Bewertung, bei der Sie beide Szenarien gleichzeitig betrachten und einen Vergleich anstellen. Sie können eine Regel anwenden. Wenn Sie der Ansicht sind, dass das zweite Szenario einen höheren Schadensersatz verdient, sollten Sie ihm einen höheren Dollarwert zuschreiben.

Nach praktisch einhelliger Auffassung sollte der Schadensersatz in beiden Situationen gleich hoch sein: Er wird für die bleibende Invalidität

geleistet – wieso sollte dann der Ort, an dem die Verletzung geschah, einen Unterschied machen? Die gemeinsame Beurteilung der beiden Szenarien gab Ihnen Gelegenheit, die moralischen Prinzipien zu prüfen, nach denen Sie die für die Opferentschädigung relevanten Faktoren bewerten. Für die meisten Menschen ist der Ort, an dem eine Körperverletzung stattfand, keiner dieser Faktoren. Wie in anderen Situationen, die einen expliziten Vergleich verlangen, war dies ein langsamer Denkprozess unter Beteiligung von System 2.

Die Psychologen Dale Miller und Cathy McFarland, die die beiden Szenarien ursprünglich entwickelten, legten sie verschiedenen Probanden zur getrennten Beurteilung vor. In ihrem Experiment (mit Between-Subjects-Design, also zwei Probandengruppen mit jeweils eigener experimenteller Bedingung) wurde jedem Teilnehmer nur ein Szenario dargeboten, und er wurde aufgefordert, diesem einen Dollarwert zuzuschreiben. Wie Sie bestimmt schon ahnen, fanden sie heraus, dass dem Opfer eine viel höhere Schadensersatzsumme zugesprochen wurde, wenn es in einem Laden angeschossen wurde, den es nur selten aufsuchte, nicht in seinem Stammgeschäft. Betroffenheit (ein Gefühl, das eng mit Reue verwandt ist) ist ein kontrafaktisches Gefühl, das hervorgerufen wird, weil einem der Gedanke »wenn er doch nur in seinem Stammgeschäft eingekauft hätte …« spontan einfällt. Die uns vertrauten System-1-Mechanismen der Ersetzung und der Intensitätsabstimmung übersetzen die Stärke der emotionalen Reaktion auf die Fallgeschichte in eine monetäre Skala, was dazu führt, dass sich die zuerkannten Schadensersatzsummen stark voneinander unterscheiden.

Der Vergleich der beiden Experimente enthüllt einen scharfen Gegensatz. Praktisch jeder, dem beide Szenarien zusammen dargeboten wurden (in einem Within-Subject-Design, bei dem sämtliche experimentelle Bedingungen mit einer Probandengruppe durchgespielt werden), pflichtet dem Grundsatz bei, dass Betroffenheit kein legitimes Kriterium der Beurteilung sei. Leider kommt dieser Grundsatz nur dann zum Tragen, wenn die beiden Szenarien zusammen präsentiert werden, aber so verhält es sich im Leben für gewöhnlich nicht. Der Normalfall ist hier vielmehr der Between-Subjects-Modus, bei dem gegensätzliche Alternativen, die einen zum Umdenken veranlassen könnten, nicht vorhanden sind und bei dem außerdem gemäß der WYSIATI-Regel nur die aktuell verfügbaren Informationen berücksichtigt werden. Infolgedessen bestimmen die

Überzeugungen, die man bei der moralischen Beurteilung eines Sachverhalts unterstützt, nicht unbedingt die emotionalen Reaktionen. Und die moralischen Intuitionen, die einem in verschiedenen Situationen einfallen, sind nicht in sich widerspruchsfrei. Die Diskrepanz zwischen der getrennten und der gemeinsamen Bewertung des Raubüberfall-Szenarios gehört zu einer umfangreichen Kategorie von Phänomenen der Urteils- und Entscheidungsumkehr.[2] Die ersten Präferenzumkehrungen *(preference reversals)* wurden in den frühen 1970er-Jahren entdeckt, und im Lauf der Jahre wurden viele weitere Beispiele für andere Varianten beschrieben.

## Eine Herausforderung für die Ökonomik

Das Phänomen der Präferenzumkehr hat einen wichtigen Platz in der Geschichte des Gesprächs zwischen Psychologen und Ökonomen.[3] Sarah Lichtenstein und Paul Slovic, zwei Psychologen, die zur gleichen Zeit wie Amos an der Universität Michigan promovierten, beschrieben Präferenzumkehrungen, die große Beachtung fanden. Sie führten ein Experiment über Präferenzen bei Wetten durch, das ich in einer leicht vereinfachten Version wiedergebe:

> Ihnen wird eine Wahl zwischen zwei Wetten angeboten, die bei einem
> Roulette mit 36 Fächern gespielt werden sollen.
>
> Wette A: Eine Chance von 11:36, 160 Dollar zu gewinnen, und eine Chance von 25:36, 15 Dollar zu verlieren.
>
> Wette B: Eine Chance von 35:36, 40 Dollar zu gewinnen, und eine Chance von 1:36, 10 Dollar zu verlieren.

Man bittet Sie, zwischen einer sicheren und einer riskanteren Wette zu wählen: ein fast sicherer Gewinn eines geringen Betrages oder eine niedrige Chance, einen viel größeren Betrag zu gewinnen, und eine hohe Wahrscheinlichkeit, einen Verlust zu machen. Man geht lieber auf Nummer sicher, und daher ist B eindeutig die beliebtere Wahl.

Betrachten wir jetzt jede Wette getrennt: Wenn Sie Besitzer dieser Wette wären, was wäre dann der niedrigste Preis, zu dem Sie sie verkaufen würden? Bedenken Sie, dass Sie mit niemandem verhandeln – Ihre Aufgabe ist es, den niedrigsten Preis festzusetzen, zu dem Sie wirklich bereit wären, die Wette zu veräußern. Versuchen Sie es. Vielleicht stellen Sie fest, dass der mögliche Gewinn bei dieser Aufgabe sehr markant (salient) ist und dass Ihr Urteil über den Preis dieser Wette in diesem Wert verankert ist. Die

Ergebnisse bestätigen diese Vermutung, und der Verkaufspreis ist für Wette A höher als für Wette B. Dies ist eine Präferenzumkehr: Menschen entscheiden sich eher für B als für A, aber wenn sie sich vorstellen, nur eine der beiden zu besitzen, messen sie A einen höheren Wert bei als B. Wie in den Raubüberfall-Szenarien ereignet sich die Präferenzumkehr deshalb, weil die gemeinsame Beurteilung die Aufmerksamkeit auf einen Aspekt der Situation lenkt – die Tatsache, dass Wette A viel weniger sicher ist als Wette B –, der bei der getrennten Bewertung weniger auffällig war. Die Merkmale, die den Unterschied zwischen den Optionsbewertungen in der Einzelbeurteilung verursachten – die Betroffenheit darüber, dass das Opfer sich im »falschen« Lebensmittelgeschäft aufhielt, und die Ankerung im möglichen Gewinn –, werden unterdrückt oder irrelevant, wenn die Optionen gemeinsam bewertet werden. Die emotionalen Reaktionen von System 1 bestimmen mit viel größerer Wahrscheinlichkeit die Einzelbewertung; der Vergleich, der bei der gemeinsamen Beurteilung stattfindet, geht immer mit einer gründlicheren und anstrengenderen Bewertung einher, die System 2 in Anspruch nimmt.

Die Präferenzumkehr lässt sich durch ein Experiment mit Within-Subject-Design bestätigen, bei dem die Probanden Preise für beide Serien als Teil einer langen Liste festlegen und zwischen ihnen wählen sollen. Die Teilnehmer sind sich der Inkonsistenz nicht bewusst, und wenn sie damit konfrontiert werden, können ihre Reaktionen sehr unterhaltsam sein. Ein von Sarah Lichtenstein im Jahr 1968 geführtes Interview mit einem Teilnehmer des Experiments ist ein bleibender Klassiker auf diesem Gebiet. Die Experimentatorin unterhält sich ausführlich mit einem verunsicherten Probanden, der sich für jeweils eine von zwei Wetten entscheidet, aber dann bereit ist, Geld dafür zu bezahlen, die Option, für die er sich gerade entschieden hatte, gegen jene einzutauschen, die er gerade verworfen hatte – und dies immer wieder.[4]

Rationale *Econs* wären sicherlich nicht anfällig für Präferenzumkehrungen, und das Phänomen war daher eine Herausforderung für das Modell vom rationalen Agenten und für die Wirtschaftstheorie, die auf diesem Modell aufbaut. Die Herausforderung hätte ignoriert werden können, aber sie wurde es nicht. Einige Jahre nachdem erstmals über das Phänomen der Präferenzumkehr berichtet wurde, veröffentlichten zwei angesehene Ökonomen, David Grether und Charles Plott, einen Aufsatz in dem renommierten Fachjournal *American*

*Economic Review*, in dem sie von ihren eigenen Studien über das von Lichtenstein und Slovic beschriebene Phänomen berichteten.[5] Dies war vermutlich das erste Studienergebnis von Experimentalpsychologen, das jemals die Aufmerksamkeit von Ökonomen auf sich zog. Der einleitende Absatz des Artikels von Grether und Plott klang für einen wissenschaftlichen Beitrag ungewöhnlich dramatisch, und ihre Absicht war klar: »Die Psychologie verfügt heute über einen umfassenden Fundus an empirischen Daten und theoretischen Ressourcen, der auch für Wirtschaftswissenschaftler von Interesse sein sollte. Auf den ersten Blick scheinen die Daten mit der Präferenztheorie unvereinbar zu sein und weitreichende Folgen für die Forschungsprioritäten innerhalb der Wirtschaftswissenschaften zu haben. Dieser Aufsatz berichtet über die Ergebnisse einer Reihe von Experimenten, die die Arbeiten von Psychologen in ihrer ökonomischen Anwendbarkeit diskreditieren sollten.«

Grether und Plott listeten 13 Theorien auf, die die ursprünglichen Ergebnisse erklären konnten, und beschrieben sorgfältig konzipierte Experimente, die diese Theorien testeten. Eine ihrer Hypothesen, die Psychologen selbstverständlich als herablassend empfanden, lautete, die Ergebnisse seien darauf zurückzuführen, dass das Experiment von Psychologen ausgeführt wurde! Schließlich blieb nur eine Hypothese übrig: Die Psychologen hatten recht. Grether und Plott räumten ein, dass diese Hypothese aus der Sicht der Standardpräferenztheorie die unbefriedigendste ist, weil »sie zulässt, dass individuelle Wahlhandlungen von dem Kontext abhängen, in dem sie stattfinden« – was ein klarer Verstoß gegen die Kohärenzdoktrin war.[6]

Vielleicht glauben Sie, dieses überraschende Ergebnis hätte jede Menge zerknirschte Gewissensprüfungen bei den Ökonomen hervorgebracht, da eine Grundannahme ihrer Theorie erfolgreich infrage gestellt wurde. Aber so läuft es nicht in den Sozialwissenschaften, zu denen ja auch die Psychologie und die Volkswirtschaftslehre zählen. Theoretische Überzeugungen sind robust, und es bedarf mehr als eines widerlegenden Befundes, damit anerkannte Theorien ernsthaft angezweifelt werden.[7] Tatsächlich hatte Grether und Plotts bewundernswert unverblümter Bericht keine nennenswerten direkten Auswirkungen auf die Anschauungen von Ökonomen, vermutlich Grether und Plott selbst eingeschlossen. Er förderte jedoch die Bereitschaft der Ökonomenzunft, die psychologische Forschung

ernst zu nehmen, und hat dadurch das interdisziplinäre Gespräch stark vorangebracht.

## Kategorien

»Wie groß ist John?« Wenn John 1,52 Meter groß ist, hängt Ihre Antwort von seinem Alter ab; er ist sehr groß, wenn er sechs und sehr klein, wenn er 16 Jahre alt ist. Ihr System 1 ruft automatisch die relevante Norm ab, und die Bedeutung des Maßstabs der Körpergröße wird automatisch korrigiert. Sie sind auch in der Lage, transkategorial Intensitäten abzugleichen und die Frage zu beantworten: »Wie teuer ist ein Gericht in einem Restaurant, das Johns Körpergröße entspricht ?« Ihre Antwort hängt von Johns Alter ab: ein viel weniger teures Gericht, wenn er 16, als wenn er sechs ist. Aber betrachten Sie jetzt folgendes Beispiel:

John ist 6. Er ist 1,52 Meter groß.
Jim ist 16. Er ist 1,55 Meter groß.

Bei Einzelbewertungen besteht Einigkeit darüber, dass John sehr groß ist und Jim nicht, weil sie mit verschiedenen Normen verglichen werden. Wenn man Ihnen eine direkt vergleichende Frage stellt – »Ist John genauso groß wie Jim?« –, werden Sie mit Nein antworten. Das ist nicht weiter überraschend und ziemlich eindeutig. In anderen Situationen jedoch kann der Prozess, durch den Objekte und Ereignisse ihren eigenen Vergleichskontext schaffen, bei ernsten Angelegenheiten zu inkohärenten Entscheidungen führen.

Sie sollten nicht den Eindruck gewinnen, dass einzelne und gemeinsame Bewertungen immer inkonsistent oder dass Urteile völlig chaotisch sind. Unsere Welt wird in Kategorien zerlegt, für die wir Normen haben, wie etwa sechsjährige Jungen oder Tische. Urteile und Präferenzen sind kohärent innerhalb von Kategorien, aber potenziell inkohärent, wenn die Objekte, die beurteilt werden, verschiedenen Kategorien angehören.

Die folgenden drei Fragen liefern ein Beispiel dafür. Beantworten Sie sie:

Was mögen Sie lieber, Äpfel oder Pfirsiche?
Was mögen Sie lieber, Steak oder Eintopf?
Was mögen Sie lieber, Äpfel oder Steak?

Die erste und die zweite Frage beziehen sich auf Objekte, die zur selben Kategorie gehören, und Sie wissen sofort, was Sie lieber mögen. Außerdem hätten Sie bei Einzelbewertung (»Wie sehr mögen Sie Äpfel?« und »Wie sehr mögen Sie Pfirsiche?«) die gleiche Rangfolge aufgestellt, weil Äpfel und Pfirsiche beide die Kategorie »Obst« aufrufen. Es kommt zu keiner Präferenzumkehr, weil verschiedene Obstsorten bei der einzelnen wie bei der gemeinsamen Bewertung mit derselben Norm und implizit auch miteinander verglichen werden. Im Gegensatz zu den intrakategorialen Fragen gibt es keine stabile Antwort für den Vergleich zwischen Äpfeln und Steaks. Anders als Äpfel und Pfirsiche sind Äpfel und Steaks keine natürlichen Substitute; sie befriedigen nicht das gleiche Bedürfnis. Manchmal hat man Lust auf ein Steak und manchmal auf einen Apfel, aber man wird nur selten sagen, dass das eine genauso gut ist wie das andere.

Stellen Sie sich vor, Sie bekommen eine E-Mail von einer Organisation, der Sie im Allgemeinen vertrauen und die Sie um eine Spende bittet:

> Viele Fortpflanzungsgebiete von Delfinen sind durch Wasserverschmutzung bedroht, und man erwartet, dass dies zu einem Rückgang der Delfinpopulation führen wird. Ein aus Privatspenden finanzierter Sonderfonds wurde eingerichtet, um schadstofffreie Fortpflanzungsgebiete für Delfine zu schaffen.

Welche Assoziationen hat das in Ihnen ausgelöst? Ihnen sind Gedanken und Erinnerungen an ähnliche Spendenaktionen eingefallen, auch wenn Sie sich dessen nicht unbedingt voll und ganz bewusst waren.

Mit hoher Wahrscheinlichkeit werden Projekte zum Schutz bedrohter Arten abgerufen. Bewertungen auf der Gut-Schlecht-Dimension sind eine automatische Operation von System 1, und Sie machten sich ein ungefähres

Bild von dem Rangplatz des Delfins unter den Arten, die Ihnen einfielen. Delfine sind viel ansprechender als etwa Frettchen, Schnecken oder Karpfen – sie haben einen sehr hohen Rang in der Menge der Arten, mit denen sie spontan verglichen werden.

Die Frage, die Sie beantworten müssen, lautet nicht, ob Sie Delfine lieber mögen als Karpfen; Sie wurden gebeten, einen Wert (in Dollar) anzugeben. Vielleicht wissen Sie aus Erfahrungen mit früheren Spendenaufrufen, dass Sie nie auf solche Bitten reagieren. Stellen Sie sich trotzdem ein paar Minuten lang als jemanden vor, der bei solchen Aufrufen spendet.

Wie viele andere schwierige Fragen lässt sich auch die Abschätzung eines Dollarwertes durch Ersetzung und Intensitätsabstimmung lösen. Die Dollar-Frage ist schwierig, aber eine leichtere Frage ist mühelos verfügbar. Da Sie Delfine mögen, sind Sie vermutlich der Ansicht, dass es eine gute Sache ist, sie zu retten. Der nächste, ebenfalls automatisch erfolgende Schritt generiert eine Dollarzahl, indem er die Intensität Ihrer Zuneigung zu Delfinen in eine Spendenskala übersetzt. [8] Sie haben eine ungefähre Vorstellung von Ihrer Skala früherer Spenden für Umweltschutzprojekte, die sich von der Skala Ihrer Spenden für politische Parteien oder für das Footballteam Ihrer Universität unterscheiden mag. Sie wissen, welcher Betrag eine »sehr große« Spende für Sie wäre und was »groß«, »bescheiden« und »klein« bedeutet. Sie verfügen auch über Skalen Ihrer Einstellung zu Arten (von »mag ich sehr« bis »überhaupt nicht«). Daher können Sie Ihre Einstellung in die Dollarskala übersetzen und sich automatisch von »mag ich sehr« zu »ziemlich große Spende« und von dort zu einer Dollarzahl bewegen. Bei einer weiteren Gelegenheit tritt man mit einem anderen Aufruf an Sie heran:

> Landarbeiter, die viele Stunden der Sonne ausgesetzt sind, haben eine höhere Hautkrebsrate als der Bevölkerungsdurchschnitt. Häufige medizinische Vorsorgeuntersuchungen können das Risiko verringern. Ein Fonds wird eingerichtet, um Vorsorgeuntersuchungen für bedrohte Gruppen zu finanzieren.

Ist dies ein dringendes Problem? Welche Kategorie wurde als eine Norm aufgerufen, als Sie die Dringlichkeit beurteilten? Wenn Sie das Problem automatisch als ein Anliegen der öffentlichen Gesundheitsversorgung einstuften, waren Sie vermutlich der Meinung, dass das Hautkrebsrisiko von Landarbeitern unter diesen Anliegen nicht besonders hoch rangiert – höchstwahrscheinlich niedriger als der Rangplatz von Delfinen auf der Liste der bedrohten Arten. Als Sie Ihren Eindruck von dem relativen Stellenwert der Hautkrebsfrage in einen Dollarbetrag übersetzten, kam dabei vielleicht ein kleinerer Betrag heraus als der, den Sie zum Schutz einer bedrohten Tierart anboten. In Experimenten verbuchten die Delfine bei einer Einzelbewertung etwas höhere Spendenbeträge als die Landarbeiter.

Betrachten wir die beiden Anliegen jetzt in gemeinsamer Bewertung. Welches der beiden Schutzobjekte, Delfine oder Landarbeiter, verdient einen höheren Dollarbetrag? Die gemeinsame Bewertung unterstreicht ein Merkmal, das bei der Einzelbewertung nicht bemerkbar war, aber als entscheidend erkannt wird, sobald es bemerkt wird: Landwirte sind Menschen, Delfine nicht. Das wussten Sie selbstverständlich, aber es war für Ihr Urteil bei der Einzelbewertung nicht relevant. Die Tatsache, dass Delfine keine Menschen sind, wurde Ihnen nicht bewusst, weil alle Gedächtnisinhalte, die bei Ihnen aktiviert wurden, ebenfalls dieses Merkmal besaßen. Die Tatsache, dass Landarbeiter Menschen sind, kam Ihnen nicht in den Sinn, weil alle Probleme des öffentlichen Gesundheitswesens Menschen betreffen. Aufgrund des engen Framings der Einzelbewertung hatten Delfine einen höheren Intensitätswert, der infolge des Intensitätsabgleichs zu einem hohen Spendenbeitrag führte. Die gemeinsame Bewertung verändert die Repräsentation der Sachverhalte: Das Merkmal »Mensch oder Tier« zieht nur dann die Aufmerksamkeit auf sich, wenn die beiden zusammen dargeboten werden. Bei der gemeinsamen Bewertung zeigen Menschen eine deutliche Präferenz für die Landarbeiter und die Bereitschaft, einen erheblich höheren Beitrag für ihr gesundheitliches Wohlergehen zu leisten als für den Schutz einer liebenswerten nicht menschlichen Art. Wie bei den Wetten und der Schussverletzung bei dem Raubüberfall stimmen die Urteile bei der Einzelbewertung und bei der gemeinsamen Bewertung auch hier nicht überein.

Das folgende Beispiel einer Präferenzumkehr stammt von Christopher Hsee von der Universität Chicago. Bei den Objekten, die bewertet werden

sollen, handelt es sich um gebrauchte Musikwörterbücher.

| | Wörterbuch A | Wörterbuch B |
|---|---|---|
| Erscheinungsjahr | 1993 | 1993 |
| Zahl der Einträge | 10 000 | 20 000 |
| Zustand | wie neu | Einband zerschlissen, sonst wie neu |

Wenn die Wörterbücher einzeln beurteilt werden, wird Wörterbuch A höher bewertet, aber selbstverständlich verändert sich die Präferenz bei der gemeinsamen Bewertung. Das Ergebnis verdeutlicht Hsees Bewertbarkeitshypothese: Die Anzahl der Einträge wird bei der Einzelbewertung nicht gewichtet, weil die Zahlen als solche nicht »bewertbar« sind. Bei der gemeinsamen Bewertung dagegen ist unmittelbar offensichtlich, dass Wörterbuch B in diesem Merkmal überlegen ist, und es ist auch offensichtlich, dass die Anzahl der Einträge viel wichtiger ist als der Zustand des Einbands.[9]

## Ungerechte Umkehrungen

Wir haben guten Grund zu der Annahme, dass die Rechtsprechung auf mehreren Gebieten unter vorhersagbaren Inkohärenzen leidet. Die Belege stammen zum Teil aus Experimenten, darunter Studien über »Pseudo-Geschworenengerichte«, und zum Teil aus der Beobachtung von Mustern in Gesetzgebung, Regulierung und bei Gerichtsverfahren.

In einem Experiment sollten Pseudo-Geschworene, die aus Geschworenenlisten in Texas rekrutiert wurden, die angemessene Höhe des sogenannten »Strafschadensersatzes« in mehreren Zivilfällen beurteilen. Diese Fälle wurden jeweils paarweise präsentiert – jedes Paar bestand aus einer Schadensersatzforderung für Körperverletzung und einer Entschädigungsforderung für finanzielle Verluste. Die Pseudo-Geschworenen beurteilten zunächst eines der Szenarien, dann wurde ihnen der Fall gezeigt, der damit verbunden war, und sie wurden aufgefordert, die beiden zu vergleichen. Nachfolgend die Zusammenfassungen eines Fallpaares:

> Fall 1: Ein Kind erlitt leichte Verbrennungen, als sein Pyjama beim Spielen mit Streichhölzern Feuer fing. Die Firma, die die Pyjamas herstellt, hat diese nicht ausreichend feuerfest gemacht.

> Fall 2: Die skrupellosen Geschäfte einer Bank trugen einer anderen Bank einen Verlust von 10 Millionen Dollar ein.

Die Hälfte der Teilnehmer beurteilte zunächst Fall 1 (in Einzelbewertung), bevor sie die beiden Fälle in gemeinsamer Bewertung verglich. Bei den übrigen Teilnehmern wurde die Reihenfolge umgekehrt. Bei der Einzelbewertung erkannten die Geschworenen der betrogenen Bank eine

höhere Schadensersatzsumme zu als dem verbrannten Kind, vermutlich weil das Ausmaß des finanziellen Verlustes einen hohen Anker bereitstellte.

Doch als die Fälle zusammen betrachtet wurden, überwog das Mitgefühl für das individuelle Opfer den Ankereffekt, und die Geschworenen sprachen dem Kind jetzt eine höhere Entschädigung zu als der Bank. Über mehrere solche Fallpaare gemittelt, wurden Opfern von Körperverletzungen bei gemeinsamer statt einzelner Bewertung mehr als doppelt so hohe Schadensersatzsummen zugesprochen. Die Geschworenen, die den Fall des an Verbrennungen leidenden Kindes gesondert betrachteten, machten ein Angebot, das der Intensität ihrer Gefühle entsprach. Sie konnten nicht vorhersehen, dass die dem Kind zugesprochene Schadensersatzsumme vor dem Hintergrund einer hohen Entschädigung für ein Finanzinstitut unzureichend erscheinen würde. Bei der gemeinsamen Bewertung blieb der der Bank zugebilligte Strafschadensersatz in dem Verlust verankert, den sie erlitten hatte, aber dem verletzten Kind wurde eine höhere Entschädigung zuerkannt, worin sich die Empörung über das fahrlässige Verhalten widerspiegelt, das die Verletzung des Kindes verursachte.

Wie wir gesehen haben, werden breitere und umfangreichere Frames den Anforderungen der Rationalität eher gerecht, und die gemeinsame Bewertung ist offensichtlich breiter als die Einzelbewertung. Natürlich sollten Sie bei der gemeinsamen Bewertung auf der Hut sein, wenn jemand, der kontrolliert, was Sie zu sehen bekommen, ein persönliches Interesse an dem hat, was Sie auswählen. Verkäufer lernen sehr schnell, dass sich die Manipulation des Kontextes, in dem Kunden eine Ware sehen, tief greifend auf deren Präferenzen auswirken kann. Mit Ausnahme solcher Fälle absichtlicher Manipulation sind vermutlich Vergleichsurteile, an denen notwendigerweise System 2 beteiligt ist, eher stabil als Einzelbewertungen, in denen sich oftmals die Intensität emotionaler Reaktionen von System 1 widerspiegelt. Wir würden erwarten, dass jede Institution, der an umsichtigen Urteilen gelegen ist, bemüht sein sollte, die Richter mit einem breiten Rahmen zur Beurteilung von Einzelfällen versieht. Mit Verwunderung erfuhr ich von Cass Sunstein, dass es Geschworenen, die die Höhe eines Strafschadensersatzes festsetzen sollen, ausdrücklich verboten wird, andere Fälle zu betrachten. Im Gegensatz zu dem, was der psychologische gesunde Menschenverstand erwarten würde, fördert das Rechtssystem die Einzelbewertung.

In einer anderen Studie über Inkohärenzen im Rechtssystem verglich Sunstein die Bußgelder, die verschiedene US-Behörden, wie etwa die US-Bundesagentur für Arbeitssicherheit und die US-Umweltschutzbehörde, verhängen können. Er zog das Fazit, dass »die Bußgelder innerhalb ihrer Kategorien vollkommen sachgerecht und angemessen zu sein scheinen, zumindest in dem Sinne, dass die Bußgelder an den Schweregrad des Rechtsverstoßes gebunden sind. Bei Verstößen gegen Arbeitssicherheits- und Gesundheitsvorschriften werden die höchsten Bußgelder bei wiederholten Verstößen verhängt, die zweithöchsten bei Verstößen, die vorsätzlich begangen werden und schwerwiegend sind, und die niedrigsten bei Verstößen gegen Dokumentationspflichten.«[10] Es sollte jedoch nicht überraschen, dass die Höhe der Bußgelder zwischen den Behörden erheblich schwankte, in einer Weise, in der sich interessenpolitische und historische Erwägungen stärker widerspiegelten als globale Gerechtigkeitsanliegen. Die Höchstgrenze des Bußgeldes für einen »schwerwiegenden Verstoß« gegen Arbeitssicherheitsvorschriften beträgt 7000 Dollar, während ein Verstoß gegen das Gesetz zum Schutz von Wildvögeln mit bis zu 25 000 Dollar geahndet werden kann. Im Rahmen anderer von der jeweiligen Behörde festgesetzter Bußgelder erscheinen die Geldstrafen angemessen, aber wenn man sie miteinander vergleicht, scheinen sie nicht zusammenzupassen. Wie bei anderen Beispielen in diesem Kapitel kann man die Absurdität jedoch nur erkennen, wenn man die beiden Fälle in einem breiten Frame zusammen betrachtet. Das System der Bußgelder ist innerhalb der jeweiligen Behörden kohärent, aber global gesehen inkohärent.

## Zum Thema »Umkehrungen«

> »Die BTU-Einheiten [der Energie] sagten mir gar nichts, bis ich erfuhr, wie stark sie bei Klimaanlagen schwanken. Die gemeinsame Bewertung war entscheidend.«

»Sie sagten, dies sei eine herausragende Rede, weil Sie sie mit ihren anderen Reden verglichen. Verglichen mit den Reden anderer waren sie nach wie vor schlechter.«

»Oftmals gelangt man durch Erweiterung des Frames zu sachgerechteren Entscheidungen.«

»Wenn man Fälle isoliert betrachtet, lässt man sich vermutlich von einer emotionalen Reaktion von System 1 leiten.«

# *34. Frames und Wirklichkeit*

Italien und Frankreich bestritten das Finale der Fußballweltmeisterschaft 2006. Die nächsten beiden Sätze beschreiben beide das Ergebnis: »Italien hat gewonnen.« – »Frankreich hat verloren.« Haben diese Aussagen die gleiche Bedeutung? Die Antwort hängt zur Gänze davon ab, was man unter *Bedeutung* versteht.

Logisch gesehen, sind die beiden Beschreibungen des Spielergebnisses miteinander austauschbar, weil sie ein und denselben Weltzustand bezeichnen. Ihre Wahrheitsbedingungen sind gleich, wie die Philosophen sagen: Wenn einer dieser Sätze wahr ist, ist auch der andere wahr. So verstehen *Econs* die Welt. Ihre Überzeugungen und Präferenzen sind realitätsgebunden. Insbesondere sind die Objekte ihrer Wahlen Weltzustände, die nicht von den Wörtern beeinflusst werden, mit denen diese Zustände beschrieben werden.

Es gibt noch einen anderen Sinn von *Bedeutung*, wo »Italien hat gewonnen« und »Frankreich hat verloren« keineswegs die gleiche Bedeutung hat. In diesem Sinne ist die Bedeutung eines Satzes das, was sich in Ihrer assoziativen Maschinerie ereignet, während Sie ihn verstehen. Die beiden Sätze lösen grundverschiedene Assoziationen aus. »Italien hat gewonnen« ruft Gedanken an die italienische Mannschaft hervor und an das, was sie getan hat, um zu gewinnen. »Frankreich hat verloren« ruft Gedanken an die französische Mannschaft hervor und an das, was sie getan hat, dass sie verlor, einschließlich jenes denkwürdigen Kopfstoßes des französischen Stars Zidane gegen einen italienischen Spieler. Im Hinblick auf die Assoziationen, die sie auslösen – die jeweilige Reaktion von System 1 –, haben die beiden Sätze unterschiedliche »Bedeutungen«. Die Tatsache, dass logisch gleichwertige Aussagen verschiedene Reaktionen hervorrufen, macht es für die *Humans* unmöglich, so verlässlich rational wie die *Econs* zu sein.

## Emotionales Framing

Amos und ich benutzten den Begriff Framing-Effekte für ungerechtfertigte Einflüsse von Formulierungen auf Überzeugungen und Präferenzen.[1] Hier eines der von uns benutzten Beispiele:

> Würden Sie eine Lotterie eingehen, die eine 10-prozentige Chance, 95 Dollar zu gewinnen, und eine 90-prozentige Chance, 5 Dollar zu verlieren, bietet?

> Würden Sie 5 Dollar bezahlen, um an einer Lotterie teilzunehmen, die eine 10-prozentige Chance, 100 Dollar zu gewinnen, und eine 90-prozentige Chance, nichts zu gewinnen, bietet?

Nehmen Sie sich zunächst einen Moment Zeit, um sich selbst davon zu überzeugen, dass die beiden Probleme identisch sind. In beiden müssen Sie sich entscheiden, ob Sie eine unsichere geldwerte Aussicht *( prospect)*, die Sie entweder um 95 Dollar reicher oder um 5 Dollar ärmer machen wird, akzeptieren sollen. Jemand, dessen Präferenzen realitätsgebunden sind, würde beide Fragen gleich beantworten, aber solche Personen sind rar. Tatsächlich erhält eine Version viel mehr positive Antworten: die zweite. Ein schlechtes Ergebnis ist viel akzeptabler, wenn es als die Kosten eines Lotterieloses dargestellt wird, das keinen Gewinn erzielt hat, als wenn es nur als Verlieren einer Wette beschrieben wird. Das sollte uns nicht überraschen: *Verluste* rufen stärkere negative Gefühle hervor als *Kosten*. Wahlen sind nicht realitätsgebunden, weil System 1 nicht realitätsgebunden ist.

Das von uns konstruierte Problem wurde von dem beeinflusst, was wir von Richard Thaler gelernt hatten: Er sagte uns, dass er als Student an

seiner Tafel eine Karte befestigt hatte, auf der stand: »Kosten sind keine Verluste.« In seinem frühen Aufsatz über das Verbraucherverhalten beschrieb Thaler die Debatte darüber, ob Tankstellen erlaubt werden sollte, unterschiedliche Preise für Einkäufe zu verlangen, je nachdem, ob sie bar oder mit der Kreditkarte bezahlt wurden.[2] Die Kreditkartenlobby setzte sich konsequent dafür ein, eine Preisstaffelung zu verbieten, aber sie hatte eine Auffangposition: Die Preisstaffelung sollte, falls sie erlaubt sein würde, als Barrabatt, nicht als Kreditkartenaufschlag beschrieben werden. Ihre psychologische Intuition war ganz richtig: Verbraucher werden eher auf einen Abschlag verzichten als einen Aufschlag bezahlen. Beides mag ökonomisch äquivalent sein, aber es ist nicht emotional äquivalent. In einem eleganten Experiment verband ein Team von Neurowissenschaftlern am University College London eine Studie über Framing-Effekte mit Aufzeichnungen der Aktivität in verschiedenen Hirnarealen. Um zuverlässige Messwerte der Hirnaktivität zu erhalten, bestand das Experiment aus vielen Versuchsdurchgängen. Abbildung 14 veranschaulicht die beiden Etappen eines dieser Versuche.

Zunächst soll sich die Versuchsperson vorstellen, dass sie einen bestimmten Geldbetrag erhalten hat, in diesem Beispiel 50 Pfund. Anschließend soll die Versuchsperson zwischen einem sicheren Ergebnis und einer Glücksrad-Lotterie wählen. Wenn das Glücksrad auf Weiß stehen bleibt, »erhält« sie den gesamten Betrag; wenn es auf Schwarz stehen bleibt, geht sie leer aus. Das sichere Ergebnis ist nichts anderes als der Erwartungswert der Lotterie, in diesem Fall ein Gewinn von 20 Pfund.

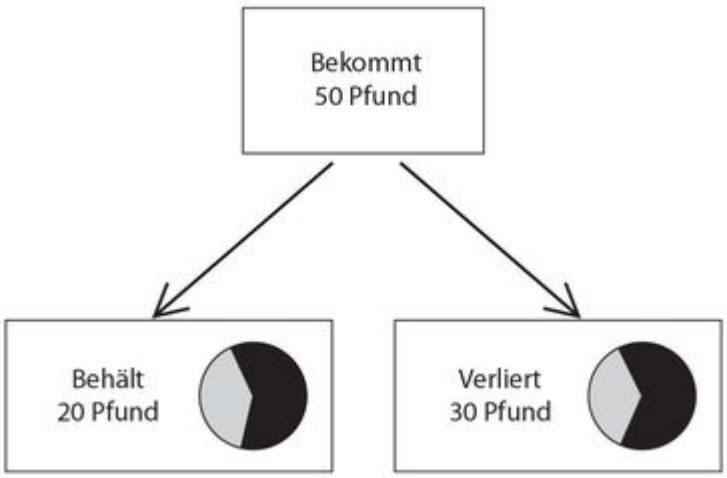


**Abbildung 14**

Wie gezeigt, lässt sich das gleiche sichere Ergebnis auf zwei verschiedene Weisen formulieren: als »behält 20 Pfund« oder als »verliert 30 Pfund«. Die objektiven Ergebnisse sind in beiden Frames genau gleich, und ein realitätstüchtiger *Econ* würde auf beide in der gleichen Weise reagieren – entweder die sichere Alternative oder, unabhängig vom Frame, das Glücksspiel wählen –, aber wir wissen bereits, dass das menschliche Gehirn nicht realitätsgebunden ist. Annäherungs- oder Vermeidungstendenzen werden durch Worte ausgelöst, und wir erwarten, dass System 1 der sicheren Option zuneigt, wenn sie als »behalten« beschrieben wird, und dieselbe Option eher ablehnt, wenn sie als »verlieren« formuliert wird.

Das Experiment bestand aus vielen Einzelaufgaben, und jedem Probanden wurden mehrere Wahlprobleme im »Behalten«- und »Verlieren«-Format dargeboten. Erwartungsgemäß zeigte jede der zwanzig Versuchspersonen einen Framing-Effekt: Sie entschieden sich mit höherer Wahrscheinlichkeit für die sichere Option in dem »Behalten«-Frame und für das Glücksspiel im »Verlieren«-Frame. Aber nicht alle Probanden waren gleich. Einige reagierten hochempfindlich auf die Darstellung des Problems. Andere trafen, unabhängig vom Frame, überwiegend stets wieder die gleiche Wahl – wie es realitätsgebundene Personen tun sollten. Die Autoren stellten eine entsprechende Rangordnung der zwanzig Probanden auf und gaben ihr die bemerkenswerte Bezeichnung »Rationalitätsindex«.

Bei jeder Entscheidung wurde die Hirnaktivität der Probanden aufgezeichnet. Später wurden die Versuchsdurchgänge (Ergebnisse der Testaufgaben) in zwei Kategorien eingeteilt:

1. Versuchsdurchgänge, bei denen die Wahl des Probanden dem Frame entsprach.
    - Bevorzugte die sichere Option in der »Behalten«-Version.
    - Bevorzugte die Lotterie in der »Verlieren«-Version.
2. Versuchsdurchgänge, bei denen die Wahl nicht dem Frame entsprach.

Die bemerkenswerten Ergebnisse veranschaulichen das Potenzial dieser neuen Disziplin Neuroökonomik – der Erforschung der Hirnaktivität bei ökonomischen Entscheidungen. Neurowissenschaftler haben Tausende derartiger Experimente durchgeführt, und sie haben gelernt, zu erwarten, dass je nach Art der Aufgabe bestimmte Hirnregionen »aufleuchten« – was auf eine erhöhte Sauerstoffversorgung hindeutet, die ihrerseits eine verstärkte neuronale Aktivität anzeigt. Verschiedene Regionen sind aktiv,

wenn ein Mensch ein visuelles Objekt wahrnimmt, sich vorstellt, einen Ball zu schießen, ein Gesicht wiedererkennt oder an ein Haus denkt. Andere Regionen leuchten auf, wenn ein Mensch emotional erregt ist, mit jemandem streitet oder sich auf die Lösung eines Problems konzentriert. Obwohl Neurowissenschaftler sorgfältig darauf achten, keine Formulierungen wie »Dieser Teil des Gehirns tut dieses und jenes« zu gebrauchen, haben sie eine Menge über die »Persönlichkeiten« verschiedener Hirnregionen gelernt, und der Stellenwert der Hirnaktivitätsanalyse bei der Interpretation psychologischer Sachverhalte hat stark zugenommen. Die Framing-Studie hat drei wichtige Befunde erbracht:

– Eine Region, die im Allgemeinen mit emotionaler Erregung assoziiert ist (die Amygdala), war mit hoher Wahrscheinlichkeit aktiv, wenn sich die Wahlen der Probanden mit dem Frame deckten. Genau das würden wir erwarten, wenn die emotional aufgeladenen Wörter »behalten« und »verlieren« eine sofortige Tendenz zur Annäherung an die sichere Option (wenn sie als Gewinn formuliert wird) oder zur Vermeidung (wenn sie als Verlust formuliert wird) auslösen. Die Amygdala reagiert sehr schnell auf emotionale Reize – und sie ist wahrscheinlich an System 1 beteiligt.
– Eine Hirnregion, die mit Konflikt und Selbstkontrolle assoziiert ist (der *Cortex cingularis anterior*), war aktiver, wenn Probanden nicht das taten, was der spontanen Tendenz entspricht – wenn sie sich für die sichere Option entschieden, obwohl diese als »verlieren« bezeichnet wurde. Es löst offensichtlich einen inneren Konflikt aus, der Neigung von System 1 zu widerstehen.
– Die »rationalsten« Probanden – diejenigen mit der geringsten Anfälligkeit für Framing-Effekte – zeigten eine verstärkte Aktivität in einer frontalen Hirnregion, die an der Verbindung von Emotion und logischem Denken bei der Entscheidungsfindung beteiligt ist. Bemerkenswerterweise waren die »rationalen« Individuen nicht jene, die die stärksten neuronalen Hinweise auf innere Konflikte zeigten. Offenbar waren diese »auserlesenen« Probanden (oft, nicht immer) realitätsgebunden, ohne nennenswerten inneren Konflikt.

Dadurch, dass diese Studie Beobachtungen tatsächlicher Wahlhandlungen mit der Aufzeichnung neuronaler Aktivität verband, verdeutlichte sie, wie

die durch ein Wort hervorgerufene Emotion in die endgültige Entscheidung »durchsickern« kann.

Ein Experiment, das Amos mit Kollegen an der Harvard Medical School durchführte, ist das klassische Beispiel für emotionales Framing. Die teilnehmenden Ärzte erhielten statistische Daten über die Ergebnisse zweier Behandlungsmethoden für Lungenkrebs: Operation und Bestrahlung. Die Fünf-Jahres-Überlebensrate spricht eindeutig für eine Operation, aber auf kurze Sicht ist eine Operation riskanter als die Bestrahlung. Die eine Hälfte der Teilnehmer las statistische Angaben über die Überlebensraten, die andere erhielt die gleichen Informationen über die Sterblichkeitsraten. Die beiden Beschreibungen der kurzfristigen Ergebnisse einer Operation lauteten:

> Die Ein-Monats-Überlebensrate liegt bei 90 Prozent.
> Im ersten Monat beträgt die Sterblichkeitsrate 10 Prozent.

Sie kennen bereits die Ergebnisse: Die Operation war beim ersten Frame (84 Prozent der Ärzte entschieden sich dafür) beliebter als beim zweiten (wo 50 Prozent eine Bestrahlung vorzogen). Die logische Äquivalenz der beiden Beschreibungen ist offensichtlich, und ein realitätsgebundener Entscheider würde unabhängig von der ihm dargebotenen Version immer die gleiche Wahl treffen. Aber System 1, wie wir es kennen, ist nur selten unempfänglich für emotional aufgeladene Wörter: Sterblichkeit ist schlecht, Überleben ist gut, und eine Überlebensrate von 90 Prozent hört sich ermutigend an, während eine Sterblichkeitsrate von 10 Prozent erschreckend klingt.[3] Ein wichtiger Befund der Studie war, dass Ärzte genauso anfällig für den Framing-Effekt waren wie medizinisch ungebildete Laien (Klinikpatienten und Studenten der Betriebswirtschaftslehre). Ein Medizinstudium ist offensichtlich kein Schutz gegen die Macht des Framings.

Die »Behalten-Verlieren«-Studie und das Überleben-Sterblichkeit-Experiment unterschieden sich in einem wichtigen Punkt. Die Teilnehmer an der Neuroimaging-Studie absolvierten zahlreiche Testdurchgänge, in denen ihnen verschiedene Frames dargeboten wurden. Sie hatten

Gelegenheit, die ablenkenden Effekte der Frames zu erkennen und sich ihre Aufgabe dadurch zu vereinfachen, dass sie sich einen gemeinsamen Frame zu eigen machen, etwa indem sie den »Verlieren«-Betrag in seine »Behalten«-Entsprechung übersetzen. Nur ein intelligenter Mensch (mit einem wachsamen System 2) kann dies lernen, und die wenigen Teilnehmer, denen dieses Kunststück gelang, gehörten vermutlich zu den »rationalen« Akteuren, die von den Experimentatoren identifiziert wurden. Dagegen hatten die Ärzte, die die statistischen Angaben über die beiden Therapien im Überlebens-Frame lasen, keinen Grund zur Annahme, dass sie eine andere Wahl getroffen hätten, wenn sie die gleichen statistischen Daten im Sterblichkeits-Frame gehört hätten. Das »Reframing« ist anstrengend, und System 2 ist normalerweise faul. Sofern es keinen naheliegenden Grund dafür gibt, sich anders zu verhalten, nehmen die meisten von uns Entscheidungsprobleme passiv so hin, wie sie »eingerahmt« sind, und sie haben daher nur selten Gelegenheit, das Ausmaß zu entdecken, in dem unsere Präferenzen framegebunden und nicht realitätsgebunden sind.

## Leere Intuitionen

Amos und ich begannen unsere Diskussion des Framings mit einem Beispiel, das unter dem Namen »Problem der Asiatischen Krankheit« bekannt wurde:[4]

> Stellen Sie sich vor, die Vereinigten Staaten bereiten sich auf den Ausbruch einer ungewöhnlichen asiatischen Krankheit vor, die schätzungsweise 600 Menschenleben fordern wird. Zwei alternative Programme zur Bekämpfung der Krankheit wurden vorgeschlagen. Angenommen, die exakten wissenschaftlichen Schätzungen der Folgen der Programme lauten folgendermaßen:
>
> Programm A1 wird 200 Menschenleben retten.
>
> Bei Programm B2 werden mit einer Wahrscheinlichkeit von 1 : 3 600 Menschenleben gerettet; mit einer Wahrscheinlichkeit von 2 : 3 wird kein Menschenleben gerettet.

Die große Mehrzahl der Befragten wählten Programm A1: Sie ziehen die sichere Option der Lotterie vor.

Die Erfolge des Programms werden in einer zweiten Version anders formuliert:

Bei Programm A2 werden 400 Menschen sterben.

Bei Programm B2 besteht eine Wahrscheinlichkeit von 1 : 3, dass niemand stirbt, und eine Wahrscheinlichkeit von 2 : 3, dass 600 Menschen sterben.

Schauen Sie genau hin und vergleichen Sie die beiden Versionen: Die Folgen der Programme A1 und A2 sind identisch, ebenso die Folgen der Programme B1 und B2. Doch beim zweiten Frame entschied sich eine große Mehrheit für das Glücksspiel.

Die verschiedenen Wahlen in den beiden Frames entsprechen der Neuen Erwartungstheorie, bei der Wahlen zwischen Lotterien und sicheren Optionen unterschiedlich gelöst werden, je nachdem, ob die Ergebnisse gut oder schlecht sind. Entscheider ziehen tendenziell die sichere Option der Lotterie vor (sie sind risikoscheu), wenn die Ergebnisse positiv sind. Sie verwerfen für gewöhnlich die sichere Option und akzeptieren das Glücksspiel (sie sind risikofreudig), wenn beide Ergebnisse negativ sind. Diese Schlussfolgerungen waren für Entscheidungen zwischen Lotterien und sicheren Ergebnissen, bei denen es um Geld ging, gut abgesichert. Das Krankheitsproblem zeigt, dass dieselbe Regel gilt, wenn die Ergebnisse in geretteten oder verlorenen Menschenleben gemessen werden. Auch in diesem Kontext enthüllt das Framing-Experiment, dass risikoscheue und risikofreudige Präferenzen nicht realitätsgebunden sind. Präferenzen zwischen denselben objektiven Ergebnissen kehren sich bei unterschiedlichen Formulierungen um.

Ein Erlebnis, von dem mir Amos berichtete, gab der Geschichte etwas Betrübliches. Amos wurde eingeladen, einen Vortrag vor einer Gruppe von Fachleuten aus dem öffentlichen Gesundheitswesen zu halten – die Personen, die über Impfungen und andere Programme entscheiden. Er nutzte die Gelegenheit, um ihnen das »Problem der Asiatischen Krankheit« zu präsentieren: Die Hälfte sah die Version mit den »geretteten Menschenleben«, die anderen beantworteten die Frage nach den

»verlorenen Menschenleben«. Wie andere Menschen waren auch diese Fachleute anfällig für Framing-Effekte. Es ist einigermaßen beunruhigend, dass die Amtsträger, die Entscheidungen treffen, welche sich auf die Gesundheit jedes Einzelnen auswirken, durch eine so oberflächliche Manipulation beeinflusst werden können – aber wir müssen uns mit der Vorstellung vertraut machen, dass auch wichtige Entscheidungen von System 1 beeinflusst, wenn auch nicht bestimmt werden.

Noch beunruhigender ist das, was geschieht, wenn Menschen mit ihrer eigenen Inkonsistenz konfrontiert werden: »Sie haben entschieden, in einer Formulierung 200 Menschenleben sicher zu retten, und in der anderen Formulierung haben Sie sich entschieden, eine riskante Wette einzugehen, statt sich mit 400 Todesopfern abzufinden. Jetzt, wo Sie wissen, dass diese Wahlen inkonsistent sind, wie entscheiden Sie sich?« Die Antwort ist im Allgemeinen ein betretenes Schweigen. Die Intuitionen, die ausschlaggebend waren für die ursprüngliche Wahl, stammten von System 1 und hatten ebenso wenig wie die Präferenz für das Behalten von 20 Pfund oder die Aversion gegen das Verlieren von 30 Pfund eine moralische Grundlage. Mit Sicherheit Menschenleben zu retten ist gut, Todesfälle sind schlecht. Die meisten Menschen stellen fest, dass ihr System 2 keine eigenen moralischen Intuitionen hat, um die Frage zu beantworten.

Ich danke dem bedeutenden Wirtschaftswissenschaftler Thomas Schelling für mein Lieblingsbeispiel eines Framing-Effekts, das er in seinem Buch *Choice and Consequence* beschrieben hat.[5] Schelling schrieb sein Buch, bevor unsere Arbeit über Framing veröffentlicht wurde, und das Framing stand nicht im Mittelpunkt seines Interesses. Er berichtete über seine Erfahrungen bei einem Seminar an der Kennedy School in Harvard, in dem es um Kinderfreibeträge im Steuergesetz ging. Schelling teilte seinen Studenten mit, dass es einen einheitlichen Freibetrag für jedes Kind gebe und dass die Höhe des Freibetrags unabhängig vom Einkommen des Steuerzahlers sei. Er fragte sie nach ihrer Meinung zu dem folgenden Vorschlag:

> Sollte der Kinderfreibetrag für die Reichen größer sein als für die Armen?

Ihre eigenen Intuitionen sind sehr wahrscheinlich die gleichen wie diejenigen von Schellings Studenten: Sie fanden die Vorstellung völlig unannehmbar, die Reichen durch einen größeren Freibetrag zu begünstigen.

Anschließend wies Schelling darauf hin, dass das Steuergesetz willkürlich sei. Es unterstelle, dass eine kinderlose Familie der Normalfall sei, und vermindere die Steuerbelastung um den Freibetrag für jedes Kind. Das Steuergesetz könne natürlich mit einem anderen Normalfall umgeschrieben werden: einer Familie mit zwei Kindern. Bei dieser Formulierung würden Familien mit weniger als der Normalzahl der Kinder einen Zuschlag bezahlen. Jetzt bat Schelling seine Studenten, ihre Meinung zu einem anderen Vorschlag zu äußern:

> Sollten die kinderlosen Armen einen genauso großen Zuschlag zahlen wie die kinderlosen Reichen?

Auch hier sind Sie wahrscheinlich mit der Reaktion der Studenten auf diese Idee einverstanden, die sie genauso vehement ablehnten wie die erste. Aber Schelling zeigte seinen Studenten, dass es logisch inkonsistent war, beide Vorschläge abzulehnen. Stellen Sie beide Formulierungen nebeneinander. Die Differenz zwischen der Steuerschuld einer kinderlosen Familie und der einer Familie mit zwei Kindern wird in der ersten Version als eine Steuersenkung und in der zweiten als eine Steuererhöhung beschrieben. Wenn man will, dass die Armen in der ersten Version die gleiche (oder eine größere) Vergünstigung für Kinder erhalten wie die Reichen, dann muss man auch wollen, dass die Armen für Kinderlosigkeit mindestens den gleichen Zuschlag bezahlen wie die Reichen.

Wir sehen System 1 in seinem Element. Es erzeugt eine sofortige Antwort auf eine Frage über die Reichen und die Armen: Im Zweifel sollen die Armen begünstigt werden. Der überraschende Aspekt von Schellings Problem besteht darin, dass diese scheinbar einfache moralische Regel nicht zuverlässig funktioniert. Sie erzeugt widersprüchliche Antworten für dasselbe Problem, je nachdem, wie es formuliert wird. Und natürlich kennen Sie bereits die Frage, die als Nächstes kommt. Jetzt, wo Sie gesehen haben, dass Ihre Reaktionen auf das Problem von dem Frame beeinflusst

werden, wie lautet Ihre Antwort auf die Frage: »Wie sollte das Steuergesetz die Kinder der Reichen und der Armen behandeln?«

Auch hier werden Sie vermutlich verblüfft sein. Sie haben moralische Intuitionen über Unterschiede zwischen den Reichen und den Armen, aber diese Intuitionen hängen von einem willkürlichen Referenzpunkt ab, und sie beziehen sich nicht auf das reale Problem. Dieses Problem – die Frage nach tatsächlichen Weltzuständen – lautet, wie viel Steuern einzelne Familien zahlen, wie die Zellen in der Matrix des Steuergesetzes gefüllt werden sollten. Es gibt keine zwingenden moralischen Intuitionen, die uns bei der Lösung dieses Problems leiten könnten. Unsere moralischen Empfindungen sind an Frames gebunden, an Beschreibungen der Wirklichkeit statt an die Wirklichkeit als solche. Daraus folgt eine klare Erkenntnis über die Natur des Framings: Framing sollte nicht als ein Eingriff betrachtet werden, der eine zugrunde liegende Präferenz maskiert oder verzerrt. Zumindest in diesem Fall – und auch bei dem »Problem der Asiatischen Krankheit« und der Frage nach Operation oder Bestrahlung bei Lungenkrebs – gibt es keine zugrunde liegende Präferenz, die durch den Frame maskiert oder verzerrt würde. Unsere Präferenzen beziehen sich auf Probleme, die in einer bestimmten Weise formuliert wurden, und unsere moralischen Intuitionen beziehen sich auf Beschreibungen, nicht auf das Wesen einer Sache.

## Gute Frames

Nicht alle Frames sind gleich, und einige Frames sind eindeutig besser als alternative Beschreibungen (oder mentale Repräsentationen) des gleichen Sachverhalts. Betrachten Sie die folgenden beiden Probleme:

> Eine Frau hat zwei Theaterkarten für je 80 Dollar gekauft. Als sie im Theater eintrifft, öffnet sie ihr Portemonnaie und bemerkt, dass die Eintrittskarten nicht da sind. Kauft sie zwei weitere Karten, um das Stück zu sehen?

> Eine Frau geht ins Theater und will zwei Eintrittskarten kaufen, die je 80 Dollar kosten. Als sie im Theater eintrifft, öffnet sie ihr Portemonnaie und bemerkt zu ihrer Bestürzung, dass die 160 Dollar, mit denen sie die Karten kaufen wollte, fehlen. Sie könnte ihre Kreditkarte benutzen. Wird sie die Karten kaufen?

Befragte, denen nur eine Version dieses Problems dargeboten wird, gelangen, je nach Frame, zu unterschiedlichen Schlussfolgerungen. Die meisten glauben, dass die Frau im ersten Fallbeispiel nach Hause gehen wird, ohne sich die Vorstellung anzusehen, wenn sie die Karten verloren hat; und die meisten glauben, dass sie die Karten für die Vorstellung mit der Kreditkarte bezahlen wird, wenn sie das Geld verloren hat.

Die Erklärung sollte bereits vertraut sein – dieses Problem betrifft die mentale Buchhaltung und den Fehlschluss aus versunkenen Kosten. Die verschiedenen Frames rufen verschiedene mentale Konten auf, und die Bedeutung des Verlusts hängt von dem Konto ab, auf dem er verbucht wird. Wenn Karten für eine bestimmte Vorstellung verloren gehen, liegt es nahe, sie auf dem Konto zu verbuchen, das mit diesem Spiel verbunden ist. Die Kosten scheinen sich verdoppelt zu haben, und sie sind jetzt möglicherweise höher, als einem das Erlebnis wert ist. Bei einem Verlust

von Bargeld dagegen wird ein »allgemeines Einnahmekonto« belastet – die Theaterbesucherin ist etwas ärmer, als sie geglaubt hatte, und sie wird sich wahrscheinlich die Frage stellen, ob die kleine Verringerung des verfügbaren Einkommens ihre Entscheidung über den Kartenkauf verändern wird. Die meisten Befragten glauben das nicht.

Die Version, in der Geld verloren wurde, führt zu sachgerechteren Entscheidungen. Sie ist ein besserer Frame, weil der Verlust, selbst wenn Karten verloren gegangen sind, »versunken« ist, und versunkene Kosten sollten ignoriert werden. Die Vergangenheit ist irrelevant, und das Einzige, was zählt, sind die Optionen, die die Theaterbesucherin jetzt hat, und ihre wahrscheinlichen Folgen. Was immer sie verloren hat, die relevante Tatsache ist, dass sie jetzt weniger vermögend ist, als sie es war, bevor sie ihr Portemonnaie öffnete. Falls mich die Person, die die Karten verloren hat, um Rat fragen würde, würde ich ihr sagen: »Hätten Sie Eintrittskarten gekauft, wenn Sie die gleiche Summe Bargeld verloren hätten? Wenn ja, dann nur zu, kaufen Sie sich neue.« Breitere Frames und umfassendere Konten führen im Allgemeinen zu rationaleren Entscheidungen.

Beim nächsten Beispiel lösen zwei alternative Frames verschiedene mathematische Intuitionen aus, und eine davon ist der anderen weit überlegen. In einem Artikel mit dem Titel »The MPG Illusion«, der 2008 im Wissenschaftsjournal *Science* erschien, identifizierten die Psychologen Richard Larrick und Jack Soll einen Fall, in dem die passive Akzeptanz eines irreführenden Frames mit erheblichen Kosten und schwerwiegenden politischen Konsequenzen verbunden ist.[6] Die meisten Autokäufer führen den Kraftstoffverbrauch als einen der Faktoren an, der ihre Kaufentscheidung bestimmt; sie wissen, dass Autos mit hoher Kilometerleistung niedrigere Betriebskosten haben. Aber der Frame, der traditionellerweise in den Vereinigten Staaten benutzt wurde – Meilen pro Gallone (MPG) –, bietet eine sehr schlechte Orientierung für die Entscheidungen von Privatpersonen und Politikern. Betrachten wir zwei Autobesitzer, die ihre Kosten senken wollen:

> Adam steigt von einem Spritfresser mit 12 Meilen pro Gallone auf eine etwas weniger gefräßige Spritschleuder, die es auf 14 Meilen pro Gallone bringt.

Die umweltbewusste Beth steigt von einem Auto mit 30 Meilen pro Gallone auf ein Auto mit 40 Meilen pro Gallone um.

Angenommen, beide Fahrer legen im Lauf eines Jahres gleiche Entfernungen zurück. Wer spart durch die Umstellung mehr Kraftstoff ein? Sie werden vermutlich sofort die weitverbreitete Intuition teilen, Beths Schritt sei bedeutender als der von Adam: Sie erhöhte die Meilen pro Gallone um 10 Meilen statt nur um 2, das heißt um ein Drittel (von 30 auf 40 Meilen) statt um ein Sechstel (von 12 auf 14 Meilen). Aktivieren Sie jetzt Ihr System 2 und berechnen Sie es. Wenn die beiden Autobesitzer jeweils 10 000 Meilen fahren, wird Adam seinen Verbrauch von skandalösen 833 Gallonen auf noch immer erschütternde 714 Gallonen verringern, also 119 Gallonen einsparen. Beths Spritverbrauch wird von 333 Gallonen auf 250 Gallonen sinken, was einer Ersparnis von 83 Gallonen entspricht. Der MPG-Frame ist falsch, und er sollte durch den Gallonen-pro-Meile-Frame (oder Liter-pro-100-Kilometer-Frame, der in den meisten anderen Ländern benutzt wird) ersetzt werden. Larrick und Soll weisen darauf hin, dass die irreführenden Intuitionen, die durch den Meilen-pro-Gallone-Frame begünstigt werden, Politiker wahrscheinlich genauso täuschen werden wie Autokäufer.

Unter Präsident Obama leitet Cass Sunstein das Office of Information and Regulatory Affairs (Amt für Informationspolitik und Regulierung). Zusammen mit Richard Thaler hat Sunstein das Buch *Nudge* geschrieben, den elementaren Leitfaden für die Anwendung der Verhaltensökonomik im Bereich der Politik. Es ist kein Zufall, dass der Aufkleber mit Angaben über den Kraftstoffverbrauch und die Umweltfreundlichkeit, der ab 2013 in den Vereinigten Staaten auf jedem Neuwagen angebracht werden muss, zum ersten Mal die Gallonen pro Meile angeben wird. Leider wird die richtige Formulierung in Kleindruck erscheinen, neben der geläufigeren Meilen-pro-Gallone-Information in Großdruck, aber es ist ein Schritt in die richtige Richtung. Die fünfjährige Zeitspanne zwischen der Veröffentlichung von »The MPG Illusion« und der Umsetzung einer teilweisen Korrektur ist vermutlich ein Geschwindigkeitsrekord für eine bedeutende politische Nutzanwendung psychologischer Erkenntnisse.

Eine Anweisung über die Organspende im Fall eines tödlichen Verkehrsunfalls ist in vielen Ländern im Führerschein einer Person vermerkt. Die Formulierung dieser Anweisung ist ein weiterer Fall, in dem ein bestimmter Frame einem anderen eindeutig überlegen ist. Nur wenige Menschen würden behaupten, dass die Entscheidung darüber, ob man seine Organe spenden will oder nicht, belanglos ist, aber vieles deutet darauf hin, dass die meisten Menschen ihre Entscheidung unbedacht treffen. Die entsprechenden Daten stammen aus einem Vergleich der Organspenderraten in europäischen Staaten, die bemerkenswerte Unterschiede zwischen benachbarten und kulturell ähnlichen Ländern aufdecken.[7] Ein im Jahr 2003 veröffentlichter Artikel weist darauf hin, dass die Organspenderrate in Österreich nahe bei 100 Prozent liegt, während sie in Deutschland nur 12 Prozent beträgt, hingegen 86 Prozent in Schweden, aber nur 4 Prozent in Dänemark.

Diese gewaltigen Unterschiede sind ein Framing-Effekt, der durch das Format der entscheidenden Frage verursacht wird. Die Länder mit hoher Organspendenbereitschaft haben eine Nichteinverständnis-Klausel; Personen, die ihre Organe nicht spenden wollen, müssen ein entsprechendes Kästchen ankreuzen. Ansonsten gelten sie als bereitwillige Spender. Die Länder mit geringer Spendenbereitschaft haben einen ausdrücklichen Einverständnisvorbehalt: Man muss ein Kästchen markieren, um zu einem Spender zu werden. Das ist alles. Der beste Prädiktor für die Organspendenbereitschaft von Menschen ist die Bezeichnung der Standardoption, die gilt, ohne dass man ein Kästchen ankreuzen muss.

Anders als andere Framing-Effekte, die auf Merkmale von System 1 zurückgeführt wurden, lässt sich der Organspendeneffekt am besten mit der Faulheit von System 2 erklären. Menschen werden das Kästchen ankreuzen, wenn sie bereits beschlossen haben, was sie tun wollen. Wenn sie nicht auf die Frage vorbereitet sind, müssen sie angestrengt darüber nachdenken, ob sie das Kästchen ankreuzen wollen. Ich stelle mir ein Organspendenformular vor, in dem Menschen ein mathematisches Problem in dem Kästchen lösen sollen, das ihrer Entscheidung entspricht. Eines der Kästchen enthält das Problem 2 + 2 = ? Das Problem im anderen Kästchen ist 13 × 37 = ? Die Spendenrate würde dadurch mit Sicherheit beeinflusst werden.

Wenn die Bedeutung der Formulierung anerkannt wird, erhebt sich eine politische Frage: Welche Formulierung sollte gewählt werden? In diesem

Fall ist die Antwort einfach. Wenn man der Ansicht ist, ein großes Angebot an Spenderorganen sei im Interesse der Gesellschaft, wird man nicht neutral sein zwischen einer Formulierung, die eine fast hundertprozentige Spendenbereitschaft hervorruft, und einer anderen Formulierung, die nur 4 Prozent der Autofahrer zu Spendern macht.

Wie wir immer wieder gesehen haben, wird eine wichtige Entscheidung von einem völlig belanglosen Merkmal der Situation bestimmt. Das ist beschämend – wir würden wichtige Entscheidungen lieber anders treffen. Außerdem erleben wir selbst diese gedanklichen Prozesse ganz anders, aber es ist unbestreitbar, dass diese kognitiven Illusionen existieren.

Wir können das als ein Argument gegen die Theorie vom rationalen Agenten zählen. Eine Theorie, die diesen Namen verdient, behauptet, dass gewisse Ereignisse unmöglich sind – sie ereignen sich nicht, wenn die Theorie wahr ist. Wird ein »unmögliches« Ereignis beobachtet, gilt die Theorie als widerlegt. Theorien können noch lange Zeit, nachdem sie durch schlüssige Beweise widerlegt wurden, fortbestehen, und das Modell des rationalen Agenten hat die empirischen Belege, die wir gesehen haben, und auch viele andere Gegenbeweise überdauert.

Der Fall der Organspenden zeigt, dass die wissenschaftliche Debatte über die Frage, wie rational wir Menschen uns verhalten, weitreichende Folgen in der gesellschaftlichen Wirklichkeit hat. Ein bedeutender Unterschied zwischen denjenigen, die an das Modell des rationalen Agenten glauben, und den Skeptikern, die es bezweifeln, besteht darin, dass seine Verfechter es einfach als selbstverständlich erachten, dass die Formulierung eines Entscheidungsproblems die diesbezüglichen Präferenzen nicht determiniert. Sie interessieren sich nicht einmal dafür, das Problem zu erforschen – und so müssen wir uns oft mit suboptimalen Ergebnissen begnügen.

Skeptiker, die die Rationalität des Menschen eher als gering einstufen, überrascht das nicht. Sie haben gelernt, eine besondere Sensibilität für scheinbar belanglose Faktoren zu entwickeln, die unsere Präferenzen bestimmen – ich hoffe, dass die Leser dieses Buches die gleiche Sensibilität erworben haben.

## Zum Thema »Frames und Wirklichkeit«

»Sie werden das, was geschehen ist, positiver sehen, wenn es ihnen gelingt, das Ergebnis danach zu framen, wie viel Geld sie behalten haben, statt wie viel sie verloren haben.«

»Wir wollen das Problem reframen, indem wir den Referenzpunkt verändern. Stellen wir uns vor, es würde uns nicht gehören – wie viel wäre es uns wert?«

»Belasten Sie Ihr mentales Konto für ›allgemeine Einnahmen‹ mit dem Verlust – Sie werden sich besser fühlen!«

»Sie sollen ein Kästchen ankreuzen, um sich aus ihrer Mailingliste auszutragen. Diese Liste würde schrumpfen, wenn Sie aufgefordert würden, ein Kästchen anzukreuzen, um in die Mailingliste aufgenommen zu werden!«

# TEIL V

## Zwei Selbste

# *35. Zwei Selbste*

Der Begriff »Nutzen« beziehungsweise »Nützlichkeit« *(utility)* hat im Verlauf seiner langen Geschichte zwei verschiedene Bedeutungen gehabt. Jeremy Bentham eröffnete seine *Einführung in die Prinzipien der Moral und der Gesetzgebung* mit dem berühmten Satz: »Die Natur hat die Menschheit unter die Herrschaft zweier souveräner Gebieter – ›Unlust‹ (*pain*, Leid) und ›Lust‹ (*pleasure*, Freude) – gestellt. Es ist an ihnen allein, aufzuzeigen, was wir tun sollen, wie auch zu bestimmen, was wir tun werden.«[1] In einer umständlich formulierten Fußnote entschuldigte sich Bentham dafür, dass er das Wort »Nutzen« auf diese Erfahrungen anwandte – es sei ihm nicht gelungen, ein besseres Wort zu finden. Zur genaueren Charakterisierung von Benthams Interpretation dieses Begriffs werde ich den Terminus »erfahrener Nutzen« verwenden.

In den letzten hundert Jahren haben Ökonomen das Wort in einem anderen Sinn verwendet. So, wie Ökonomen und Entscheidungstheoretiker den Terminus benutzen, bedeutet er »Wollbarkeit« *(wantability)*[2] – das, was ich »Entscheidungsnutzen« genannt habe. In der Erwartungsnutzentheorie zum Beispiel geht es ausschließlich um die Regeln der Rationalität, nach denen der Entscheidungsnutzen beurteilt werden sollten; sie hat nichts über lustvolle Erfahrungen zu sagen. Die beiden Nutzenbegriffe decken sich selbstverständlich dann, wenn Menschen das wollen, was sie genießen, und genießen, wofür sie sich entscheiden – und diese Annahme der Deckungsgleichheit ist in der allgemeinen Vorstellung, dass ökonomische Akteure rational handeln und entscheiden, implizit enthalten. Es wird angenommen, dass rationale Akteure ihre gegenwärtigen und zukünftigen Präferenzen kennen, und sie sollen sachgerechte Entscheidungen treffen, die diese Interessen maximal fördern.

## Erfahrungsnutzen

Meine Faszination für die möglichen Diskrepanzen zwischen Erfahrungs- und Entscheidungsnutzen reicht weit zurück. Als Amos und ich noch an der Neuen Erwartungstheorie arbeiteten, formulierte ich ein Problem, das folgendermaßen lautete: Stellen Sie sich eine Person vor, die jeden Tag eine schmerzhafte Injektion bekommt. Es findet keine Gewöhnung statt; der Schmerz ist Tag für Tag der gleiche. Messen Menschen der Verringerung der Anzahl der geplanten Injektionen von 20 auf 18 den gleichen Wert bei wie einer Reduktion von sechs auf vier? Gibt es irgendeine Rechtfertigung dafür, gegebenenfalls einen Unterschied zu machen?

Ich habe keine Daten erhoben, weil das Ergebnis auf der Hand lag. Sie können es selbst überprüfen: Sie würden für die Verringerung der Anzahl der Injektionen um ein Drittel (von sechs auf vier) mehr bezahlen als für eine Reduktion um ein Zehntel (von 20 auf 18). Der Entscheidungsnutzen der Vermeidung zweier Injektionen ist im ersten Fall höher als im zweiten, und jeder wird für die erste Reduktion mehr bezahlen als für die zweite. Aber dieser Unterschied ist absurd. Wenn sich der Schmerz von Tag zu Tag nicht ändert, was sollte dann rechtfertigen, dass man der Verringerung des Gesamtschmerzes um zwei Injektionen – in Abhängigkeit von der Anzahl der vorhergehenden Injektionen – verschiedene Nutzwerte zuschreibt? In der heutigen Begrifflichkeit formuliert, führte das Problem die Vorstellung ein, der Erfahrungsnutzen lasse sich mit der Anzahl der Injektionen messen. Es legte auch den Schluss nahe, dass der Erfahrungsnutzen, zumindest in einigen Fällen, das Kriterium ist, nach dem eine Entscheidung beurteilt werden sollte. Ein Entscheider, der verschiedene Geldbeträge zahlt, um den gleichen Gewinn an Erfahrungsnutzen zu erreichen (oder sich den gleichen Verlust zu ersparen), macht einen Fehler. Vielleicht finden Sie diese Feststellung selbstverständlich, aber in der Entscheidungstheorie ist die einzige Basis dafür, eine Entscheidung als falsch zu beurteilen, die Inkonsistenz mit anderen Präferenzen. Amos und ich diskutierten das Problem, aber wir verfolgten es nicht weiter. Viele Jahre später kam ich darauf zurück.

## Erfahrung und Gedächtnis

Wie lässt sich der erfahrene Nutzen messen? Wie sollten wir Fragen beantworten wie »Wie viel Schmerzen litt Helen während des medizinischen Eingriffs ?« oder »Wie viel Vergnügen bereiteten ihr zwanzig Minuten am Strand?«. Der britische Ökonom Francis Edgeworth machte sich im 19. Jahrhundert Gedanken über dieses Thema und wartete mit der Idee eines »Hedonimeters« auf, eines imaginären Instruments, ähnlich den Geräten, die in Wetterstationen eingesetzt werden, das die Intensität der Lust beziehungsweise Unlust, die eine Person zu jedem beliebigen Zeitpunkt erlebt, messen sollte.[3]

Der Erfahrungsnutzen sollte genauso schwanken wie die Tagestemperatur oder der Luftdruck, und die Ergebnisse sollten als eine Funktion der Zeit dargestellt werden. Die Antwort auf die Frage, wie viel Schmerz oder Lust Helen während ihrer medizinischen Behandlung oder ihres Strandaufenthalts erlebte, wäre die »Fläche unter der Kurve«. Die Zeit spielt in Edgeworths Konzeption eine entscheidende Rolle. Wenn Helen vierzig statt zwanzig Minuten am Strand verbringt und ihr Vergnügen genauso intensiv bleibt, dann verdoppelt sich der gesamte Erfahrungsnutzen dieser Episode – so wie die Verdopplung der Anzahl der Injektionen einer Injektionsserie doppelt so schmerzhaft ist. Dies war Edgeworths Theorie, und wir haben heute eine recht genaue Vorstellung von den Bedingungen, unter denen sie gilt.[4]

Die Diagramme in Abbildung 15 zeigen Profile der Erfahrungen von zwei Patienten, die sich einer schmerzhaften Darmspiegelung unterzogen. Sie stammen aus einer Studie, die Don Redelmeier und ich gemeinsam entwarfen. Redelmeier, Arzt und Forscher an der Universität Toronto, führte sie Anfang der 1990er-Jahre durch.[5] Bei dieser Untersuchung werden heute routinemäßig ein Narkotikum und ein amnestisches Medikament eingesetzt, aber als unsere Daten erhoben wurden, waren diese Medikamente noch nicht so weit verbreitet. Die Patienten wurden alle sechzig Sekunden nach der Intensität der Schmerzen gefragt, die sie gerade in diesem Moment verspürten. Die hier gezeigten Daten liegen auf einer Skala, auf der null »schmerzfrei« und zehn »unerträgliche Schmerzen« bedeutet. Wie Sie

sehen können, hatten beide Patienten während der Untersuchung, die acht Minuten für Patient A und 24 Minuten für Patient B dauerte (die letzte Ablesung eines Schmerzes von null wurde nach dem Ende der Untersuchung aufgezeichnet), sehr unterschiedliche Schmerzerlebnisse. An dem Experiment nahmen insgesamt 154 Patienten teil; die kürzeste Untersuchung dauerte vier, die längste 69 Minuten.

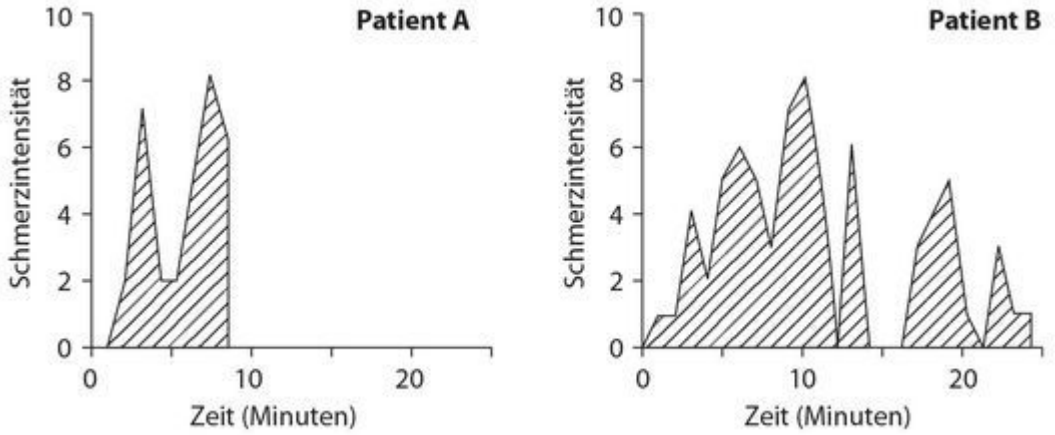


**Abbildung 15**

Betrachten Sie als Nächstes eine leichte Frage: Angenommen, die beiden Patienten benutzten die Schmerzskala in ähnlicher Weise, welcher Patient litt dann mehr Schmerzen? Das ist offensichtlich. Es besteht allgemeines Einvernehmen darüber, dass es Patient B schlechter ging. Patient B verbrachte mindestens so viel Zeit wie Patient A auf jeder Stufe der Schmerzintensität, und die »Fläche unter der Kurve« ist bei B eindeutig größer als bei A. Ausschlaggebend ist dabei natürlich die Tatsache, dass die Untersuchung von B viel länger dauerte. Ich werde die Messwerte auf der Basis von Berichten über das momentane Schmerzempfinden »Hedonimeter-Gesamtsummen« nennen.

Am Ende sollten alle Teilnehmer die Gesamtsumme der Schmerzen angeben, die sie während der Untersuchung erlebt hatten. Diese Formulierung sollte sie dazu ermuntern, an die Gesamtheit der Schmerzen zu denken, die sie angegeben hatten; auf diese Weise sollten sie die Hedonimeter-Gesamtsummen reproduzieren. Erstaunlicherweise taten die Patienten nichts dergleichen. Die statistische Analyse enthüllte zwei Befunde, die ein Muster verdeutlichen, das wir bereits bei anderen Experimenten beobachtet haben:

– Höchststand-Ende-Regel: Die globale Einstufung in der Rückschau wurde von der durchschnittlichen Schmerzintensität im schlimmsten Moment des Experiments und an seinem Ende gut vorhergesagt.
– Vernachlässigung der Dauer: Die Dauer der Untersuchung hatte nicht die geringste Auswirkung auf die Einschätzung der Schmerzsumme.

Wir können diese Regeln jetzt auf die Profile der Patienten A und B anwenden. Das schlimmste Rating (acht auf der Zehn-Punkte-Skala) war für beide Patienten gleich, aber das letzte Rating vor dem Ende der Untersuchung war sieben für Patient A und nur eins für Patient B. Daher lag der Höchststand-Ende-Mittelwert bei 7,5 für Patient A und nur bei 4,5 für Patient B. Erwartungsgemäß behielt Patient A eine viel schlechtere Erinnerung an die Episode als Patient B. Es war Pech für Patient A, dass die Untersuchung in einem schmerzhaften Moment endete und so eine unangenehme Erinnerung bei ihm zurückließ.

Wir haben jetzt ein überreiches Angebot: zwei Maße des Erfahrungsnutzens – die Hedonimeter-Gesamtsumme und die retrospektive Einschätzung –, die sich systematisch voneinander unterscheiden. Die Hedonimeter-Gesamtsummen werden von einem Beobachter aus den von einer Person berichteten momentanen Schmerzerfahrungen berechnet. Wir nennen diese Urteile »durationsgewichtet«, weil die Berechnung der »Fläche unter der Kurve« alle Momente gleich gewichtet: Zwei Minuten Schmerzen der Intensitätsstufe neun sind doppelt so schlecht wie eine Minute auf dem gleichen Schmerzniveau. Doch die Ergebnisse dieses Experiments und anderer Untersuchungen zeigen, dass die retrospektiven Einschätzungen nicht von der Dauer beeinflusst werden und zwei bestimmte Momente – den Höchststand und das Ende – viel stärker gewichten als andere. An welchem Maß sollten wir uns orientieren? Was sollte der Arzt tun? Die Entscheidung hat Auswirkungen auf die medizinische Praxis. Wir trafen folgende Feststellungen:

– Wenn es darum geht, die Schmerzerinnerung der Patienten zu verringern, könnte die Senkung der höchsten Schmerzintensität wichtiger sein als die Reduktion der Untersuchungslänge. Aus dem gleichen Grund mag eine allmähliche Linderung einer abrupten Linderung vorzuziehen sein, falls die Patienten die Untersuchung in besserer Erinnerung behalten, wenn die Schmerzen an deren Ende vergleichsweise milde sind.

– Wenn es darum geht, die tatsächlich erlebte Schmerzmenge zu reduzieren, mag die schnelle Durchführung der Untersuchung die geeignete Maßnahme sein, auch wenn dies die höchste Schmerzintensität steigert und Patienten eine schreckliche Erinnerung an die Untersuchung behalten werden.

Welches dieser beiden Ziele erscheint Ihnen vorrangiger? Ich habe keine wissenschaftliche Erhebung durchgeführt, aber mein Eindruck ist, dass eine große Mehrheit eine Verringerung der im Gedächtnis gespeicherten Schmerzerfahrung vorziehen wird. Ich finde es hilfreich, dieses Dilemma als einen Interessenkonflikt zwischen zwei Selbsten (die *nicht* den vertrauten Systemen entsprechen) aufzufassen. Das »erlebende Selbst« beantwortet die Frage: »Tut es jetzt weh?«; das »erinnernde Selbst« beschäftigt sich hingegen mit der Frage: »Wie war es im Großen und Ganzen?« Erinnerungen sind alles, was uns von unseren Lebenserfahrungen bleibt, und die einzige Perspektive, die wir uns zu eigen machen können, wenn wir über unser Leben nachdenken, ist daher die des erinnernden Selbst.

Der Kommentar eines Zuhörers nach einem Vortrag, den ich gehalten hatte, verdeutlicht die Schwierigkeit, Erinnerungen von Erfahrungen zu unterscheiden. Er erzählte, er habe hingebungsvoll einer langen Symphonie auf einer CD gelauscht, die zum Ende hin zerkratzt gewesen sei, was entsetzliche Geräusche erzeugt habe. Und er sagte, dieser missratene Schluss habe »ihm das ganze Erlebnis verdorben«. Aber das stimmte nicht, nicht das Erlebnis als solches wurde ruiniert, sondern nur die Erinnerung daran. Das erlebende Selbst hatte eine Erfahrung gemacht, die fast zur Gänze positiv war, und der missratene Schluss konnte dies nicht ungeschehen machen, weil die Erfahrung bereits stattgefunden hatte. Mein Fragesteller hatte die gesamte Episode als ungenügend beurteilt, weil sie sehr schlecht ausgegangen war, aber diese Beurteilung ignorierte faktisch vierzig Minuten reinsten Musikgenusses. Zählt die tatsächliche Erfahrung nichts?

Die Verwechslung der Erinnerung mit der tatsächlichen Erfahrung ist eine zwingende kognitive Illusion – und es ist diese Ersetzung, die uns glauben macht, eine vergangene Erfahrung sei von Grund auf negativ gewesen. Das erlebende Selbst hat keine Stimme. Das erinnernde Selbst irrt sich manchmal, aber es ist dasjenige, das Buch führt und bestimmt, was wir aus dem Leben lernen, und es ist auch dasjenige, das Entscheidungen trifft.

Die Vergangenheit lehrt uns, die Qualitäten unserer zukünftigen Erinnerungen, nicht unbedingt unserer zukünftigen Erfahrungen, zu maximieren. Dies ist die Tyrannei des erinnernden Selbst.

## Welches Selbst sollte zählen?

Um die Entscheidungsmacht des erinnernden Selbst zu veranschaulichen, entwarfen meine Kollegen und ich ein Experiment, bei dem wir eine milde Form von Folter anwandten, die ich »Kalte-Hand-Situation« nennen werde (der hässliche Fachterminus lautet »Kältedrucktest«). Die Probanden sollen ihre Hand bis zum Handgelenk in schmerzhaft kaltes Wasser eintauchen, bis sie aufgefordert werden, sie herauszunehmen, und ihnen ein warmes Handtuch angeboten wird. Die Versuchspersonen in unserem Experiment benutzten ihre freie Hand, um Pfeile auf einer Tastatur zu steuern und so ein kontinuierliches Protokoll ihrer Schmerzempfindungen zu liefern – eine direkte Kommunikation ihres erlebenden Selbst. Wir wählten eine Temperatur, die mittlere, aber erträgliche Schmerzen verursachte: die freiwilligen Teilnehmer durften natürlich jederzeit ihre Hand aus dem kalten Wasser herausnehmen, aber niemand tat dies.

Jeder Proband machte zwei Kalte-Hand-Episoden durch:

> Die kurze Episode bestand aus einem sechzig Sekunden langen Eintauchen in Wasser mit einer Temperatur von 14 Grad Celsius, was als unangenehm, aber nicht unerträglich kalt erlebt wurde. Am Ende der sechzig Sekunden wies der Experimentator den Teilnehmer an, seine Hand aus dem Wasser herauszunehmen, und bot ihm ein warmes Handtuch an.

> Die lange Episode dauerte neunzig Sekunden. Ihre ersten sechzig Sekunden waren identisch mit der kurzen Episode. Nach den sechzig Sekunden sagte der Experimentator nichts. Stattdessen öffnete er ein Ventil, worauf geringfügig wärmeres Wasser in die Wanne lief. Während der zusätzlichen dreißig Sekunden stieg die Temperatur des Wassers um etwa ein Grad,

gerade genug für die meisten Probanden, um eine leichte Abnahme der Schmerzintensität zu empfinden.

Wir sagten unseren Probanden, dass sie drei Kalte-Hand-Versuche absolvieren würden – tatsächlich machten sie nur die kurze und die lange Episode durch, jede mit einer anderen Hand. Zwischen den beiden Versuchen lagen sieben Minuten. Sieben Minuten nach dem zweiten Versuch gab man den Teilnehmern eine Wahlmöglichkeit in Bezug auf den dritten Versuch. Man sagte ihnen, eine ihrer Erfahrungen würde wiederholt und sie könnten frei entscheiden, ob sie den Versuch mit ihrer linken oder ihrer rechten Hand wiederholen wollten.[6] Die Hälfte der Teilnehmer hatte den kurzen Versuch mit der linken Hand gemacht, die andere Hälfte mit der rechten; die eine Hälfte hatte zuerst den kurzen Versuch gemacht, die andere hatte mit dem langen begonnen und so weiter. Es handelte sich um ein sorgfältig kontrolliertes Experiment.

Das Experiment sollte einen Interessenkonflikt zwischen dem erlebenden und dem erinnernden Selbst sowie zwischen dem Erfahrungs- und dem Entscheidungsnutzen erzeugen. Aus der Sicht des erlebenden Selbst war der lange Versuch offensichtlich schlimmer. Wir erwarteten, dass das erinnernde Selbst eine andere Meinung hätte. Die Höchststand-Ende-Regel sagt vorher, dass der kurze Versuch in negativerer Erinnerung bleiben wird als der längere, und die Vernachlässigung der Dauer sagt vorher, dass die Differenz zwischen neunzig und sechzig Sekunden Schmerzerleben ignoriert wird. Daher gingen wir davon aus, dass die Probanden den langen Versuch in besserer (oder weniger schlechter) Erinnerung behalten hatten und sich dafür entscheiden würden, ihn zu wiederholen. Und das taten sie auch. Ganze 80 Prozent der Teilnehmer, die berichteten, ihre Schmerzen hätten in der Schlussphase des längeren Versuchs nachgelassen, entschieden sich, diesen zu wiederholen; sie erklärten sich also bereit, in dem vorweggenommenen dritten Versuch dreißig Sekunden lang unnötige Schmerzen zu erdulden.

Die Probanden, die die lange Episode vorzogen, waren keine Masochisten und entschieden sich nicht absichtlich dafür, sich der schlimmeren Erfahrung auszusetzen; sie machten einfach einen Fehler. Hätten wir sie gefragt: »Was ziehen Sie vor: ein neunzig Sekunden langes Eintauchen oder nur den ersten Teil davon?«, dann hätten sie mit Sicherheit

die zweite Option gewählt. Aber wir benutzten nicht diese Worte, und die Probanden taten das, was für sie selbstverständlich war: Sie beschlossen, die Episode zu wiederholen, an die sie sich mit weniger Unlust erinnerten. Die Probanden wussten recht genau, welche der beiden Reizexpositionen länger dauerte – wir fragten sie –, aber sie nutzten dieses Wissen nicht. Ihre Entscheidung folgte einer einfachen Regel intuitiver Wahl: Wähle die Option aus, die du am meisten magst beziehungsweise am wenigsten nicht magst. Gedächtnisregeln beeinflussten, wie stark ihre Abneigung gegen die beiden Optionen war, was wiederum ihre Wahl bestimmte. Wie das von mir konzipierte Injektionsproblem offenbarte auch das Kalte-Hand-Experiment eine Diskrepanz zwischen Entscheidungs- und Erfahrungsnutzen.

Die Präferenzen, die wir bei diesem Experiment beobachteten, sind ein weiteres Beispiel für den Weniger-ist-mehr-Effekt, dem wir bei früheren Gelegenheiten begegnet sind. Eine davon war Christopher Hsees Studie, bei der das Hinzufügen von Teilen zu einem 24-teiligen Geschirrservice den Gesamtwert verringerte, weil einige der zusätzlichen Geschirrteile beschädigt waren. Ein weiteres Beispiel war Linda, die Aktivistin, die mit höherer Wahrscheinlichkeit als eine feministische Bankkassiererin denn als eine Bankkassiererin identifiziert wird. Die Ähnlichkeit ist kein Zufall. Das gleiche »Betriebsmerkmal« von System 1 erklärt alle drei Situationen: System 1 repräsentiert Mengen durch Mittelwerte, Normen und Prototypen, nicht durch Summen. Jede Kalte-Hand-Episode besteht aus einer Menge von Momenten, die das erinnernde Selbst als einen prototypischen Moment abspeichert. Dies führt zu einem Konflikt. Für einen objektiven Beobachter, der die Episode auf der Basis der Berichte des erlebenden Selbst beurteilt, zählt die »Fläche unter der Kurve«, die den Schmerz im Zeitablauf integriert; sie besitzt die Eigenschaft einer Summe. Dagegen ist die Erinnerung, die das erinnernde Selbst im Gedächtnis hinterlegt, ein repräsentativer Moment, der stark von dem Höchststand und dem Ende beeinflusst wird.

Selbstverständlich hätte die Evolution das Gedächtnis der Tiere so gestalten können, dass es Integrale speichert, wie es dies in manchen Fällen sicherlich tut. Für ein Eichhörnchen ist es wichtig, die Gesamtmenge der Futtervorräte, die es angelegt hat, zu »kennen«. Eine Repräsentation der Durchschnittsgröße der Nüsse wäre kein guter Ersatz. Doch die Gesamtsumme an erlebter Unlust oder Lust im Zeitablauf mag biologisch weniger relevant sein. Wir wissen zum Beispiel, dass Ratten eine

Vernachlässigung der Dauer von Lust- und Schmerzerfahrungen zeigen. In einem Experiment wurden Ratten regelmäßig einer Folge von Reizen ausgesetzt, bei der das Aufleuchten einer Lampe die baldige Verabreichung eines Stromschlags ankündigte. Die Ratten lernten schnell, das Licht zu fürchten, und die Intensität ihrer Furcht konnte mit mehreren physiologischen Reaktionen gemessen werden. Der wichtigste Befund war, dass sich die Dauer des Schocks nicht nennenswert oder gar nicht auf die Furcht auswirkte – allein die Schmerzintensität des Reizes zählte.[7]

Andere klassische Studien zeigten, dass die elektrische Stimulation spezifischer Areale im Rattengehirn (und entsprechender Areale im menschlichen Gehirn) ein intensives Lustgefühl erzeugt, in einigen Fällen so intensiv, dass Ratten, die ihr Gehirn durch Drücken eines Hebels stimulieren können, verhungern, weil sie keine Pause zur Nahrungsaufnahme einlegen. Die lustvolle elektrische Stimulation kann in Spannungsstößen unterschiedlicher Stärke und Dauer verabreicht werden. Auch hier kommt es allein auf die Intensität an. Bis zu einem gewissen Punkt scheint die Ausweitung der Dauer eines stimulierenden elektrischen Impulses das Verlangen des Tieres danach nicht zu steigern.[8] Die Regeln, nach denen das erinnernde Selbst des Menschen funktioniert, haben eine lange evolutionäre Geschichte.

## Biologie kontra Rationalität

Die nützlichste Erkenntnis aus dem Injektionsproblem, das mich vor Jahren beschäftigte, bestand darin, dass der Erfahrungsnutzen einer Reihe gleich schmerzhafter Injektionen durch bloßes Zählen der Injektionen gemessen werden kann. Wenn alle Injektionen gleich unangenehm sind, dann sind zwanzig davon doppelt so unangenehm wie zehn, und eine Verringerung von zwanzig auf achtzehn ist genauso nützlich wie eine von sechs auf vier. Wenn der Entscheidungsnutzen nicht dem Erfahrungsnutzen entspricht, dann stimmt etwas mit der Entscheidung nicht. Die gleiche Logik zeigte sich beim Kalte-Hand-Experiment: Eine Schmerzepisode, die neunzig Sekunden dauert, ist schlimmer als die ersten sechzig Sekunden dieser Episode. Wenn Menschen sich bereitwillig entscheiden, die längere Episode über sich ergehen zu lassen, dann stimmt etwas mit ihrer Entscheidung nicht. Bei dem Injektionsproblem war die Diskrepanz zwischen der Entscheidung und der Erfahrung auf die abnehmende Empfindlichkeit zurückzuführen: Die Differenz zwischen achtzehn und zwanzig ist weniger eindrucksvoll, und sie scheint weniger wert zu sein als die Differenz zwischen sechs und vier Injektionen. In dem Kalte-Hand-Experiment spiegeln sich in dem Fehler zwei Prinzipien der Gedächtnisbildung wider: Vernachlässigung der Dauer und Höchststand-Ende-Regel. Die Mechanismen sind unterschiedlich, aber das Ergebnis ist das gleiche: eine Entscheidung, die nicht richtig auf die Erfahrung abgestimmt ist.

Entscheidungen, die nicht die bestmögliche Erfahrung hervorbringen, und Fehlprognosen zukünftiger Gefühle – beides sind schlechte Neuigkeiten für diejenigen, die an die Rationalität des menschlichen Entscheidungsverhaltens glauben. Die Kalte-Hand-Studie zeigte, dass wir nicht voll und ganz darauf vertrauen können, dass sich in unseren Präferenzen unsere Interessen widerspiegeln, selbst wenn sie auf persönlichen Erfahrungen beruhen und die Erinnerung an diese Erfahrung innerhalb der letzten Viertelstunde abgespeichert wurde! Präferenzen und Entscheidungen werden von Erinnerungen geprägt, und die Erinnerungen können falsch sein. Die empirischen Befunde erschüttern die Annahme, Menschen hätten konstante Präferenzen und wüssten, wie sie deren Nutzwert maximieren können, was ein Eckpfeiler des Modells vom

rationalen Agenten ist. In unseren Intellekt ist eine Inkonsistenz eingebaut. Wir haben starke Präferenzen bezüglich der Dauer unserer Unlust- und Lusterfahrungen. Wir wollen, dass Unlust schnell vergeht und Lust andauert. Aber unser Gedächtnis, eine Funktion von System 1, ist evolutionär dafür ausgelegt, den intensivsten Moment (den »Gipfel«) einer schmerzhaften oder lustvollen Episode und die Gefühle am Ende der Episode abzuspeichern. Ein Gedächtnis, das die Dauer vernachlässigt, ist unserer Vorliebe für lange lustvolle Erlebnisse und kurze schmerzhafte Episoden nicht dienlich.

## Zum Thema »Zwei Selbste«

> »Du beurteilst deine gescheiterte Ehe allein aus der Perspektive des erinnernden Selbst. Eine Scheidung gleicht einer Symphonie mit einem dissonanten Ausklang – die Tatsache, dass sie schlecht endete, bedeutet nicht, dass alles nur schlecht gewesen ist.«

> »Dies ist ein schlimmer Fall von Vernachlässigung der Dauer. Sie gewichten die positiven und die negativen Seiten Ihrer Erfahrung gleich stark, obwohl das Positive zehnmal so lange dauerte wie das Negative.«

# *36. Das Leben als eine Geschichte*

Als ich begann, mich wissenschaftlich mit der Frage zu beschäftigen, wie man subjektives Erleben misst, sah ich Verdis Oper *La Traviata.* Bekannt für ihre hinreißende Musik, ist sie auch eine bewegende Liebesgeschichte zwischen einem jungen Adligen und Violetta, einer Frau aus der Halbwelt. Der Vater des jungen Mannes sucht Violetta auf und überredet sie dazu, ihren Geliebten aufzugeben, um die Ehre der Familie und die Heiratschancen der Schwester des jungen Mannes zu schützen. In einem Akt höchster Selbstaufopferung gibt Violetta vor, auf den Mann, den sie über alles liebt, verzichten zu wollen. Bald darauf flammt ihre Schwindsucht (die im 19. Jahrhundert gebräuchliche Bezeichnung für Tuberkulose) auf. Im letzten Akt liegt Violetta auf dem Sterbebett, umgeben von ein paar Freunden. Als ihr Geliebter von ihrem Zustand erfährt, eilt er nach Paris, um sie zu sehen. Sobald sie die Nachricht hört, flößen ihr Hoffnung und Freude neuen Lebensmut ein; trotzdem verschlechtert sich ihr Zustand rapide.

Ganz egal, wie oft man die Oper gesehen hat, packen einen immer wieder die Spannung und die Furcht des Augenblicks: Wird der junge Liebhaber rechtzeitig eintreffen? Man spürt, dass es für ihn enorm wichtig ist, seine Geliebte noch einmal zu sehen, bevor sie stirbt. Natürlich kommt er rechtzeitig; einige wunderbare Liebesduette werden gesungen, und nach zehnminütigem herrlichem Gesang segnet Violetta das Zeitliche.

Auf dem Nachhauseweg von der Oper fragte ich mich: Weshalb sind uns diese zehn Minuten so wichtig? Ich erkannte sehr schnell, dass mir die Länge von Violettas Leben völlig egal war. Wenn man mir gesagt hätte, dass sie mit 27 Jahren starb, nicht mit 28, wie ich glaubte, hätte mich die Nachricht, dass sie ein Jahr Lebensglück verpasst hatte, überhaupt nicht berührt. Aber die Möglichkeit, die letzten zehn Minuten zu versäumen, war von großer Bedeutung. Außerdem hätten sich die Gefühle, die die Wiedervereinigung der Geliebten in mir auslösten, nicht verändert, wenn ich erfahren hätte, dass sie statt zehn Minuten eine Woche miteinander verbracht hätten. Doch wenn der Liebhaber zu spät gekommen wäre, wäre *La Traviata* eine ganz andere Geschichte gewesen. Eine Geschichte dreht

sich um bedeutende Ereignisse und denkwürdige Momente, nicht um das Vergehen der Zeit. In einer Geschichte ist die Vernachlässigung der Dauer normal, und der Schluss definiert oftmals ihren Charakter. In der Art und Weise, eine Erzählung zu gestalten, und in den Erinnerungen an Darmspiegelungen, Urlaube und Filme kommen die gleichen Kernmerkmale zum Vorschein. Das erinnernde Selbst komponiert Geschichten und bewahrt sie auf, um in Zukunft darauf zurückgreifen zu können.

Nicht nur in der Oper stellen wir uns das Leben als eine Geschichte vor und wünschen uns, dass es ein gutes Ende nimmt. Wenn wir von dem Tod einer Frau erfahren, die viele Jahre lang mit ihrer Tochter zerstritten war, wollen wir wissen, ob sie sich vor ihrem Tod noch versöhnt haben. Nicht nur die Gefühle der Tochter sind uns wichtig – wir wollen auch die Erzählung des Lebens der Mutter verbessern. Die Sorge um andere Menschen manifestiert sich oftmals in dem Interesse an der Qualität ihrer Geschichten, nicht an ihren Gefühlen. Tatsächlich können uns sogar Ereignisse, die die Geschichten von bereits verstorbenen Menschen verändern, zutiefst ergreifen. Wir empfinden Mitleid für einen Mann, der in dem Glauben starb, seine Frau habe ihn geliebt, wenn wir erfahren, dass sie viele Jahre lang einen Geliebten hatte und nur wegen des Geldes bei ihrem Mann blieb.[1] Wir bemitleiden den Ehemann, obwohl er ein glückliches Leben gelebt hat. Wir empfinden die Demütigung einer Wissenschaftlerin, die eine bedeutende Entdeckung machte, welche sich nach ihrem Tod als falsch herausstellte, obwohl sie diese Schmach nicht erlebte. Am wichtigsten aber ist natürlich, dass wir uns alle intensiv für die Erzählung unseres eigenen Lebens interessieren und unbedingt wollen, dass es eine gute Geschichte mit einem anständigen Helden wird.

Der Psychologe Ed Diener und seine Studenten fragten sich, ob die Vernachlässigung der Dauer und die Höchststand-Ende-Regel auch die Bewertung ganzer Biografien bestimmt. Sie benutzten eine kurze Lebensbeschreibung einer fiktiven Figur namens Jen, einer unverheirateten Frau ohne Kinder, die bei einem Verkehrsunfall auf der Stelle und schmerzlos starb. In einer Version von Jens Geschichte war sie ihr ganzes Leben hindurch (das entweder dreißig oder sechzig Jahre dauerte) überaus glücklich; sie liebte ihre Arbeit, machte oft Urlaub, verbrachte Zeit mit ihren Freunden und ihren Hobbys. In einer anderen Version lebte Jen fünf Jahre länger, und sie starb jetzt entweder mit 35 oder mit 65. Die

zusätzlichen Jahre wurden als angenehm, wenn auch weniger glücklich als die Zeit davor beschrieben. Nachdem jeder Teilnehmer eine kurze Biografie von Jen gelesen hatte, beantwortete er zwei Fragen: »Wie erstrebenswert erscheint Ihnen Jens Leben, aufs Ganze gesehen?« und »Wie viel Glück oder Unglück hat Jen Ihrer Meinung nach insgesamt in ihrem Leben erlebt?«.

Die Ergebnisse lieferten eindeutige Belege sowohl für die Vernachlässigung der Dauer wie für den Höchststand-Ende-Effekt. In einem Experiment mit Between-Subjects-Design (verschiedenen Probanden werden verschiedene Bedingungen dargeboten) wirkte sich die Verdopplung der Dauer von Jens Lebenszeit in keiner Weise auf die Bewertung aus, wie erstrebenswert ihr Leben sei, oder auf Urteile über das gesamte Lebensglück von Jen. Ihr Leben wurde eindeutig durch einen prototypischen Zeitabschnitt, nicht als eine Folge von Zeitabschnitten repräsentiert. Infolgedessen war ihre Gesamtzufriedenheit die Zufriedenheit eines typischen Zeitraums in ihrem Leben, nicht die Summe (oder das Integral) der Zufriedenheit über die Dauer ihres Lebens.

Ausgehend von dieser Idee, fanden Diener und seine Studenten wie zu erwarten auch einen Weniger-ist-mehr-Effekt, einen starken Hinweis darauf, dass eine Summe durch einen Durchschnitt (Prototyp) ersetzt wurde. Die Hinzufügung fünf leidlich glücklicher Jahre zu einem sehr glücklichen Leben führte zu deutlich schlechteren Bewertungen der Gesamtzufriedenheit dieses Lebens.

Auf meine Anregung hin erhoben sie auch Daten über die Auswirkungen der zusätzlichen fünf Jahre im Rahmen eines Within-Subject-Experiments: Jeder Proband gab beide Urteile unmittelbar nacheinander ab. Ungeachtet meiner langen Erfahrung mit Urteilsfehlern glaubte ich nicht, dass vernünftige Menschen behaupten würden, die Hinzufügung fünf leidlich glücklicher Jahre zu einem Leben werde dessen Bilanz erheblich verschlechtern. Ich irrte mich. Die Intuition, wonach die enttäuschenden fünf zusätzlichen Jahre die gesamte Lebensbilanz verschlechterten, war übermächtig.

Das Muster der Urteile schien so absurd zu sein, dass Diener und seine Studenten anfangs glaubten, es spiegele die Torheit der jungen Menschen wider, die an ihren Experimenten teilnahmen. Aber das Muster änderte sich nicht, als die Eltern und ältere Freunde der Studenten dieselben Fragen beantworteten. Bei der intuitiven Bewertung ganzer Lebensläufe sowie

kurzer Episoden kommt es auf Höchststände und Endzustände an, jedoch nicht auf die Dauer.[2]

Die Schmerzen der Geburtswehen und der Nutzen von Ferien werden immer als Einwände gegen die Idee der Vernachlässigung der Dauer hervorgebracht: Wir alle teilen die Intuition, dass es weit schlimmer ist, wenn die Geburtswehen 24 statt sechs Stunden dauern, und dass sechs Tage an einem schönen Urlaubsort besser sind als drei. In diesen Situationen scheint die Dauer von Belang zu sein, aber dies ist nur deshalb der Fall, weil die Qualität des Endes sich mit der Länge der Episode verändert. Die Mutter ist nach 24 Stunden ausgelaugter und hilfloser als nach sechs Stunden; der Urlauber ist nach sechs Tagen erquickter und ausgeruhter als nach drei Tagen. Das, was bei der intuitiven Bewertung solcher Episoden wirklich zählt, ist die fortschreitende Verschlechterung oder Verbesserung der andauernden Erfahrung und wie sich die Person am Ende fühlt.

## Amnestischer Urlaub

Betrachten Sie folgende Urlaubsvarianten. Wollen Sie lieber eine entspannende Woche an dem vertrauten Strand verbringen, zu dem Sie letztes Jahr fuhren? Oder wollen Sie Ihren Gedächtnisspeicher anreichern? Verschiedene Wirtschaftszweige sind entstanden, um auf diese alternativen Bedürfnisse einzugehen: Urlaubsorte bieten erholsame Entspannung; der Tourismus hilft Menschen, Geschichten zu entwerfen und Erinnerungen zu sammeln. Das fieberhafte Fotografieren vieler Touristen deutet darauf hin, dass das Speichern von Erinnerungen oftmals ein wichtiges Ziel ist und sowohl die Urlaubspläne als auch die Urlaubserfahrung prägt. Der Fotograf betrachtet die Szene nicht als einen Moment, der genossen werden sollte, sondern als eine zu gestaltende zukünftige Erinnerung. Bilder mögen für das erinnernde Selbst nützlich sein – auch wenn wir sie nur selten sehr lange oder so oft betrachten, wie wir erwartet haben, oder auch gar nicht –, aber Fotografieren ist für das erlebende Selbst eines Touristen nicht unbedingt der beste Weg, einen schönen Anblick zu genießen.

In vielen Fällen beurteilen wir touristische Aufenthalte nach der Geschichte und den Erinnerungen, die wir zu speichern hoffen. Das Wort »unvergesslich« wird oftmals dazu benutzt, Urlaubshöhepunkte zu beschreiben, und es enthüllt ausdrücklich das Ziel der Erfahrung. In anderen Situationen – hier fällt einem die Liebe ein – verändert die Erklärung, dass der gegenwärtige Moment unvergessen bleiben wird, auch wenn das nicht immer zutreffend ist, den Charakter dieses Moments. Eine das Selbstbewusstsein stärkende unvergessliche Erfahrung erhält ein Gewicht und eine Bedeutung, die sie andernfalls nicht haben würde.

Ed Diener und sein Team lieferten Belege dafür, dass die Urlaubsziele von dem erinnernden Selbst ausgesucht werden. Sie baten Studenten, Tagebuch zu führen und ihre Erfahrungen während der Semesterferien im Frühjahr täglich zu beurteilen. Die Studenten gaben auch ein globales Urteil über die Ferien ab, nachdem sie vorbei waren. Schließlich gaben sie an, ob sie noch einmal den gleichen Urlaub machen würden oder nicht. Die statistische Auswertung ergab, dass die zukünftigen Urlaubsabsichten voll und ganz von der abschließenden Bewertung bestimmt wurden – auch wenn diese Beurteilung nicht genau mit der Qualität der Erfahrungen, die in den

Tagebüchern beschrieben wurden, übereinstimmte. Wie beim Kalte-Hand-Experiment *entscheiden* wir – ob richtig oder falsch – *anhand der Erinnerungen*, wenn wir uns fragen, ob wir eine Erfahrung wiederholen wollen oder nicht.

Ein Gedankenexperiment über Ihren nächsten Urlaub erlaubt es Ihnen, Ihre Einstellung zu Ihrem erlebenden Selbst zu beobachten.

> Am Ende des Urlaubs werden alle Bilder und Videos zerstört. Außerdem trinken Sie einen Zaubertrank, der all Ihre Urlaubserinnerungen löschen wird.

> Wie würde sich diese Aussicht auf Ihre Urlaubspläne auswirken? Wie viel wären Sie bereit dafür zu zahlen, verglichen mit einem normal erinnerbaren Urlaub?

Obwohl ich die Reaktionen auf dieses Szenario nicht wissenschaftlich erforscht habe, ist mein Eindruck aus Gesprächen, dass das Löschen der Erinnerungen den Wert von Erfahrungen erheblich vermindert. In einigen Fällen behandeln sich Menschen selbst so, wie sie einen anderen Menschen mit Gedächtnisverlust behandeln würden; sie entscheiden sich, ihre »Lustbilanz« dadurch zu maximieren, dass sie an einen Ort zurückkehren, wo sie sich früher schon einmal wohlgefühlt haben. Einige Leute sagen allerdings, sie würden sich gar nicht erst die Mühe machen, überhaupt in den Urlaub zu fahren. Damit geben sie zu erkennen, dass ihnen einzig an ihrem erinnernden Selbst gelegen ist und dass ihnen ihr an Gedächtnisverlust leidendes erlebendes Selbst weniger bedeutet als ein amnestischer Fremder. Viele weisen darauf hin, dass sie weder sich selbst noch einen anderen Amnestiker losschicken würden, um Berge zu erklimmen oder durch einen Dschungel zu marschieren – weil diese Erlebnisse in der Echtzeit überwiegend unangenehm sind und ihren Wert aus der Erwartung erhalten, dass sowohl die Mühen als auch die Freude darüber, das Ziel zu erreichen, unvergesslich sein werden.

Nun zu einem weiteren Gedankenexperiment: Stellen Sie sich vor, Ihnen steht eine schmerzhafte Operation bevor, in der Sie bei vollem Bewusstsein bleiben. Man sagt Ihnen, dass Sie vor Schmerzen schreien und den Chirurgen anflehen werden, aufzuhören. Man verspricht Ihnen jedoch ein amnestisches Medikament, das jegliche Erinnerung an die Episode löschen wird. Wie beurteilen Sie diese Aussicht? Auch hier ist es nach meinen informellen Beobachtungen so, dass die meisten Menschen den Schmerzerfahrungen ihres erlebenden Selbst bemerkenswert gleichgültig gegenüberstehen. Einige sagen, es wäre ihnen völlig egal. Andere teilen meine Einstellung: Ich empfinde zwar Mitleid mit meinem leidenden Selbst, aber nicht mehr Mitleid, als ich für einen Fremden empfinden würde, der Schmerzen leidet. So seltsam es auch erscheinen mag, ich bin mein erinnerndes Selbst, und das erlebende Selbst, das mein Leben lebt, ist für mich wie ein Fremder.

## Zum Thema »Leben als Geschichte«

»Er bemüht sich verzweifelt, die Erzählung von einem Leben in Rechtschaffenheit zu schützen, die durch den jüngsten Vorfall bedroht wird.«

»Die Mühe, die er für einen One-Night-Stand aufzuwenden bereit war, ist ein Zeichen der totalen Vernachlässigung der Dauer.«

»Sie scheinen Ihren ganzen Urlaub für die Bildung von Erinnerungen zu nutzen. Vielleicht sollten Sie die Kamera beiseitelegen und den Augenblick genießen, auch wenn er nicht besonders erinnerungswert ist?«

»Sie ist Alzheimer-Patientin. Sie kann aus ihrem Leben keine Erzählung mehr machen, aber ihr erlebendes Selbst ist noch immer empfänglich für Schönheit und Freundlichkeit.«

# *37. Erlebtes Wohlbefinden*

Als ich mich vor etwa 15 Jahren für die Erforschung des psychischen Wohlbefindens zu interessieren begann, fand ich schnell heraus, dass fast alles, was über dieses Thema bekannt war, auf den Antworten von Millionen von Menschen auf geringfügige Variationen einer Erhebungsfrage beruhte, die im Allgemeinen als Maß der Lebenszufriedenheit anerkannt wurde. Die Frage richtet sich eindeutig an Ihr erinnerndes Selbst, das dazu aufgefordert wird, über Ihr Leben nachzudenken:

> Wie zufrieden sind Sie gegenwärtig, alles in allem, mit Ihrem Leben?[1]

Da ich über die Erforschung verzerrter Erinnerungen an Darmspiegelungen und schmerzhaft kalte Hände zum Thema Wohlbefinden gekommen war, bezweifelte ich naturgemäß, dass die globale Lebenszufriedenheit ein zuverlässiges Maß des Wohlbefindens sei. Da sich das erinnernde Selbst in meinen Experimenten nicht als ein guter Zeuge erwiesen hatte, konzentrierte ich mich auf das Wohlbefinden des erlebenden Selbst. Ich behauptete, »Helen war im Monat März glücklich« sei eine sinnvolle Aussage, wenn

> sie die meiste Zeit mit Aktivitäten verbrachte, die sie lieber fortsetzen als beenden würde, wenig Zeit in Situationen, denen sie gern entfliehen wollte, und – sehr wichtig, denn das Leben ist kurz – nicht zu viel Zeit in einem neutralen Zustand, in dem ihr mehr oder minder alles egal wäre.

Es gibt viele verschiedene Erfahrungen, die wir lieber fortsetzen als beenden würden, und dazu gehören sowohl geistige als auch körperliche Freuden. Eines der Beispiele für eine Situation, die Helen meines Erachtens gern fortsetzen würde, ist die völlige Versenkung in eine Aufgabe, die Mihaly Csikszentmihalyi »Flow« nennt – ein Zustand, den manche Künstler in ihren kreativen Momenten erleben und den viele andere Menschen erreichen, wenn sie gefesselt sind von einem Film, einem Buch oder einem Kreuzworträtsel: In keiner dieser Situationen will man unterbrochen werden. Ich hatte auch Erinnerungen an eine glückliche frühe Kindheit, in der ich immer weinte, wenn meine Mutter mich von meinem Spielzeug wegriss, um mit mir im Park spazieren zu gehen, und abermals weinte, wenn sie mich von der Schaukel und der Rutsche wegnahm. Der Widerwille gegen jegliche Störung war ein Anzeichen dafür, dass ich sowohl mit meinem Spielzeug als auch mit den Schaukeln Spaß hatte.

Ich schlug vor, Helens objektive Zufriedenheit exakt zu messen, so wie wir die Schmerzerfahrung der beiden Darmspiegelungspatienten gemessen hatten, indem wir ein Profil ihres Wohlbefindens in aufeinanderfolgenden Momenten ihres Lebens erstellten. Darin folgte ich Edgeworths alter Hedonimeter-Methode. In meiner anfänglichen Begeisterung für diese Methode neigte ich dazu, Helens erinnerndes Selbst als einen fehleranfälligen Zeugen des tatsächlichen Wohlbefindens ihres erlebenden Selbst abzutun. Ich vermutete, dass diese Auffassung zu extrem war, was sich auch bewahrheitete, aber es war ein guter Anfang.

## Erlebtes Wohlbefinden

Ich stellte ein Dream-Team zusammen, dem drei weitere Psychologen mit unterschiedlicher Spezialisierung und ein Ökonom angehörten, und wir begannen, ein Maß für das Wohlbefinden des erlebenden Selbst zu entwickeln.[2] Eine kontinuierliche Protokollierung der Erfahrung war leider unmöglich – eine Person kann nicht normal leben, während sie fortwährend ihr emotionales Wohlergehen dokumentiert. Die beste Alternative war die Erfahrungsstichprobe, eine von Csikszentmihalyi konzipierte Methode. Dieses Verfahren wurde seit seinen ersten Anwendungen erheblich weiterentwickelt. Bei der Umsetzung der Erfahrungsstichprobe verfährt man heute so, dass das Handy einer Testperson programmiert wird, an zufällig ausgewählten Zeitpunkten während des Tages zu piepsen oder zu vibrieren. Anschließend präsentiert das Gerät ein kurzes Menü von Fragen darüber, was der Befragte gemacht hat und wer gerade bei ihm war, als er unterbrochen wurde. Der Testperson werden auch Ratingskalen gezeigt, mit deren Hilfe sie die Intensität verschiedener Gefühle einstufen soll: Zufriedenheit, Anspannung, Wut, Sorge, Engagement (Wohlbefinden durch Arbeitsbeanspruchung), körperlicher Schmerz und andere.[3]

Die Erfahrungsstichprobe ist teuer und beschwerlich (wenn auch weniger belastend, als die meisten Menschen zunächst erwarten; die Beantwortung der Fragen geht sehr schnell). Eine praktischere Alternative wurde dringend benötigt, also entwickelten wir eine Methode, die wir *Day Reconstruction Method* (DRM – »Verfahren zur Rekonstruktion des Tagesablaufs«) nannten. Wir hofften, sie würde den Ergebnissen der Erfahrungsstichprobe nahekommen und zusätzliche Informationen darüber liefern, wie Menschen ihre Zeit verbringen. [4] Die Teilnehmer (in den frühen Studien ausnahmslos Frauen) wurden zu einer zweistündigen Sitzung eingeladen. Wir baten sie zunächst, den Vortag innerlich noch einmal detailliert nachzuvollziehen und ihn wie Szenen eines Films in Episoden zu zerlegen. Später beantworteten sie Menüs von Fragen über jede Episode, basierend auf der Methode der Erfahrungsstichprobe. Aus einer Liste wählten sie Aktivitäten aus, denen sie nachgingen, und gaben an, welcher Aktivität sie die meiste Aufmerksamkeit schenkten. Sie listeten auch die Personen auf, mit denen sie Kontakt gehabt

hatten, und stuften die Intensität mehrerer Gefühle auf separaten Skalen von null bis sechs ein (null = das Gefühl fehlt; sechs = die höchste Intensität des Gefühls). Unsere Methode stützte sich auf Befunde, die belegten, dass Menschen, die in der Lage sind, eine vergangene Situation im Detail abzurufen, auch die damit verbundenen Gefühle, ja sogar die früheren körperlichen Anzeichen emotionaler Zustände nacherleben konnten.[5]

Wir nahmen an, dass unsere Probanden sich das Gefühl eines prototypischen Moments der Episode recht genau ins Gedächtnis zurückrufen könnten. Mehrere Vergleiche mit der Erfahrungsstichprobe bestätigten die Gültigkeit der DRM. Weil die Probanden auch die Zeiten angaben, zu denen Episoden begannen und endeten, konnten wir ein nach der Dauer gewichtetes Maß ihres Wohlbefindens während des gesamten Wachtages berechnen. In unserem summarischen Maß der täglichen Befindlichkeit zählten längere Episoden mehr als kurze. Unser Fragebogen umfasste auch Maße der Lebenszufriedenheit, die wir als die Zufriedenheit des erinnernden Selbst interpretierten. Wir benutzten die DRM, um die Bestimmungsfaktoren sowohl des emotionalen Wohlbefindens als auch der Lebenszufriedenheit bei mehreren Tausend Frauen in den Vereinigten Staaten, Frankreich und Dänemark zu erforschen.

Die Erfahrung eines Moments oder einer Episode lässt sich nicht so ohne Weiteres mit einem einzigen Messwert der Zufriedenheit erfassen. Es gibt etliche Spielarten positiver Gefühle, wie etwa Liebe, Freude, Engagement, Hoffnung, Heiterkeit und viele mehr. Auch von negativen Emotionen gibt es zahlreiche Varianten, etwa Zorn, Scham, Niedergeschlagenheit und Einsamkeit. Obgleich positive und negative Emotionen gleichzeitig existieren, lassen sich die meisten Momente des Lebens letztlich als positiv oder negativ einstufen. Wir konnten unangenehme Episoden identifizieren, indem wir die Ratings positiver und negativer Adjektive miteinander verglichen. Eine Episode bezeichneten wir als unangenehm, wenn einem negativen Gefühl ein höheres Rating zugeschrieben wurde als allen positiven Gefühlen. Wir fanden heraus, dass sich Amerikanerinnen etwa 19 Prozent ihrer Zeit seelisch unwohl fühlten, etwas mehr als Französinnen (16 Prozent) und Däninnen (14 Prozent).

Wir nannten den Prozentsatz der Tageszeit, den sich eine Person unwohl fühlt, den U-Index.[6] Eine Person, die sich zum Beispiel vier Stunden ihrer täglichen Wachzeit von 16 Stunden unwohl fühlte, hätte einen U-Index von

25 Prozent. Das Reizvolle am U-Index ist die Tatsache, dass er nicht auf einer Ratingskala, sondern auf einer objektiven Messung von Zeitspannen basiert. Wenn der U-Index für eine Population von 20 auf 18 Prozent fällt, kann man daraus folgern, dass die Gesamtzeit, die die Population in einem Zustand emotionalen Unbehagens oder Leidens verbrachte, um ein Zehntel zurückgegangen ist.

Ein bemerkenswerter Befund war das Ausmaß der Ungleichheit in der Verteilung emotionaler Verstimmung.[7] Etwa die Hälfte unserer Probanden gab an, den ganzen Tag lang keine Episode seelischen Missbehagens zu erleben. Andererseits verbrachte eine beachtliche Minderheit einen Großteil des Tages in einer stark gedrückten Stimmungslage. Offenbar entfällt der größte Teil des seelischen Leids auf einen kleinen Prozentsatz der Bevölkerung – ob aufgrund einer körperlichen oder psychischen Erkrankung, einer schwermütigen Veranlagung oder wegen persönlicher Schicksalsschläge und Tragödien.

Auch für Aktivitäten lässt sich ein U-Index berechnen. So können wir zum Beispiel den Prozentsatz der Zeit messen, den Menschen beim Pendeln zur Arbeit, am Arbeitsplatz oder im Kontakt mit Eltern, Partnern oder Kindern in einem negativen emotionalen Zustand verbringen. Für tausend Amerikanerinnen einer Stadt im Mittleren Westen betrug der U-Index 29 Prozent für die morgendliche Fahrt zur Arbeit, 27 Prozent für die Zeit am Arbeitsplatz, 24 Prozent für die Kinderbetreuung, 18 Prozent für die Hausarbeit, 12 Prozent für soziale Kontakte, 12 Prozent für das Fernsehen und 5 Prozent für Sex. Der U-Index war an Wochentagen um etwa 6 Prozent höher als an Wochenenden, vor allem deshalb, weil Menschen an Wochenenden weniger Zeit mit Aktivitäten verbringen, die ihnen unangenehm sind, und nicht an der Anspannung und dem Stress leiden, die mit der Berufstätigkeit verbunden sind. Die größte Überraschung war das emotionale Erleben der Beschäftigung mit den eigenen Kindern, die für Amerikanerinnen etwas weniger angenehm war, als Hausarbeit zu verrichten. Hier fanden wir einen der wenigen Gegensätze zwischen Französinnen und Amerikanerinnen: Französinnen verbringen weniger Zeit mit ihren Kindern, genießen diese Zeit aber mehr, vielleicht weil sie Zugang zu mehr Kinderbetreuungsangeboten haben und nachmittags weniger Zeit damit verbringen, ihre Kinder zu verschiedenen Aktivitäten zu fahren.

Die Stimmungslage einer Person ist zu jedem beliebigen Zeitpunkt von ihrem Temperament und ihrer Gesamtzufriedenheit abhängig, aber das

emotionale Wohlbefinden schwankt auch erheblich im Lauf eines Tages und einer Woche. Die augenblickliche Stimmung hängt in erster Linie von der gegenwärtigen Situation ab. Das emotionale Befinden am Arbeitsplatz zum Beispiel bleibt weitgehend unberührt von den Faktoren, die die allgemeine Arbeitszufriedenheit beeinflussen, wie etwa Zusatzleistungen und Status. Wichtiger sind situative Faktoren wie etwa die Gelegenheit zum geselligen Umgang mit Kollegen, eine laute Geräuschkulisse, Zeitdruck (eine bedeutende Ursache negativer Stimmungen) und die direkte Anwesenheit eines Vorgesetzten (in unserer ersten Studie der einzige Zustand, der schlimmer war, als allein zu sein). Die Aufmerksamkeit ist ausschlaggebend. Unser emotionaler Zustand wird durch das bestimmt, worauf wir unsere Aufmerksamkeit richten, und wir konzentrieren uns normalerweise auf unsere gegenwärtige Aktivität und unser unmittelbares Umfeld. Es gibt Ausnahmen, wo die Qualität der subjektiven Erfahrung von wiederkehrenden Gedanken statt von den Ereignissen des Augenblicks bestimmt wird. Wenn wir glücklich verliebt sind, sind wir vielleicht bestens gelaunt, auch wenn wir in einem Stau stecken, und wenn wir traurig sind, bleibt unsere Stimmung vielleicht auch dann gedrückt, wenn wir einen witzigen Film ansehen. Doch unter normalen Umständen bereitet uns das, was gerade geschieht, Lust oder Unlust, sofern wir unsere Aufmerksamkeit darauf richten. Essen zum Beispiel bereitet uns nur dann Lust, wenn wir es bewusst genießen. Wir stellten fest, dass Französinnen und Amerikanerinnen für das Essen ungefähr genauso viel Zeit aufwenden, aber für die Französinnen stand es mit doppelt so hoher Wahrscheinlichkeit im Zentrum der Aufmerksamkeit wie für Amerikanerinnen. Die Amerikanerinnen kombinierten das Essen viel eher mit anderen Aktivitäten, und ihr Essensgenuss war entsprechend verdünnt.

Diese Beobachtungen haben Konsequenzen für Einzelpersonen wie für die Gesellschaft. Die Nutzung der Zeit ist einer der Lebensbereiche, über den Menschen eine gewisse Kontrolle besitzen. Nur wenige Menschen können sich allein durch Willensentschluss ein sonnigeres Gemüt zulegen, aber einige können ihr Leben vielleicht so einrichten, dass sie weniger Zeit mit Pendeln und mehr Zeit damit verbringen, mit Menschen, die sie mögen, Dinge zu tun, die ihnen Spaß machen. Die Gefühle, die mit verschiedenen Aktivitäten verbunden sind, legen nahe, dass man sein subjektives Wohlbefinden auch dadurch steigern kann, dass man von passiven Freizeitbeschäftigungen wie Fernsehen auf aktivere Formen wie soziale

Kontakte und Sport umsteigt. Gesamtgesellschaftlich gesehen sind eine verbesserte Verkehrsinfrastruktur für Berufspendler, die Ausweitung des Angebots an Kinderbetreuungseinrichtungen für berufstätige Frauen und mehr Seniorentreffs möglicherweise relativ effiziente Möglichkeiten zur Reduktion des U-Index der Gesellschaft – selbst eine Verringerung von 1 Prozent wäre ein großer Erfolg, da sie Millionen von Stunden an vermiedenem emotionalem Missbehagen bedeutet. Kombinierte nationale Erhebungen über die Nutzung der verfügbaren Zeit und das erlebte Wohlbefinden können der Gesellschaftspolitik viele wichtige Anregungen geben. Der Ökonom in unserem Team, Alan Krueger, war der Erste, der Elemente dieser Methode in die nationalen Statistiken einführte.

Das erlebte Wohlbefinden wird heute routinemäßig im Rahmen groß angelegter nationaler Befragungen in den Vereinigten Staaten, Kanada und Europa erfasst, und der Gallup World Poll hat diese Erfassung auf Millionen von Befragten in den Vereinigten Staaten und über 150 weiteren Ländern ausgeweitet.[8] Bei den Erhebungen werden die Teilnehmer gefragt, welche Gefühle sie während des Vortags erlebten, wenn auch weniger detailliert als beim DRM. Die riesigen Stichproben erlauben extrem feinmaschige Analysen, die die Bedeutung situativer Faktoren, körperlicher Gesundheit und sozialer Kontakte für das erlebte Wohlbefinden bestätigt haben. Wenn ich Kopfweh habe, fühle ich mich elend, und der zweitbeste Prädiktor für die tagesbezogene Stimmungslage ist der Kontakt zu Freunden oder Verwandten an diesem Tag. Es ist nur leicht übertrieben, zu sagen, dass Glück die Erfahrung des Zusammenseins mit Menschen ist, die einen lieben und die man liebt.

Die Gallup-Daten erlauben einen Vergleich zwischen zwei Aspekten des Wohlergehens:

– Das Wohlbefinden, das Menschen erleben, während sie ihr Leben leben
– Das Urteil, das sie fällen, wenn sie ihr Leben bewerten

Gallup misst die Beurteilung der Lebensqualität mit einer *Cantril Self-Anchoring Strive Scale* genannten Frage:

> Bitte stellen Sie sich eine Leiter mit zehn Sprossen vor, wobei null ganz unten und zehn an der Spitze ist. Das obere Ende der

> Leiter stellt das bestmögliche Leben für Sie dar, das untere Ende steht für das schlimmstmögliche Leben für Sie. Auf welcher Sprosse der Leiter stehen Sie nach Ihrer persönlichen Einschätzung zum gegenwärtigen Moment?

Einige Aspekte des Lebens wirken sich stärker auf die Beurteilung des eigenen Lebens als auf die unmittelbare Erfahrung des Lebens aus. Der Bildungsabschluss ist ein Beispiel. Ein höherer Bildungsstand geht mit einer höheren Bewertung des eigenen Lebens einher, aber nicht mit einem höheren erlebten Wohlbefinden. Tatsächlich berichten die Höherqualifizierten, zumindest in den Vereinigten Staaten, über größeren Stress. Andererseits wirkt sich ein schlechter Gesundheitszustand viel negativer auf das erlebte Wohlbefinden als auf die Lebensbewertung aus. Das Zusammenleben mit Kindern erlegt ebenfalls erhebliche Kosten in der Währung der täglichen Gefühle auf – Eltern berichten häufig von Stress und Ärger, aber die negativen Auswirkungen auf die Lebensbewertung sind viel geringer. Religiosität hat einen vergleichsweise stärkeren günstigen Effekt auf die Stimmungslage wie auch auf die Stressreduktion als auf die Lebensbewertung. Erstaunlicherweise vermindert Religiosität jedoch nicht Gefühle von Niedergeschlagenheit oder Sorge.

Die Auswertung von über 450 000 Antworten auf den *Gallup-Healthways Well-Being Index*, eine tägliche Befragung von tausend Amerikanern, liefert eine erstaunlich klare Antwort auf eine der am häufigsten gestellten Fragen der Glücksforschung: Macht Geld glücklich?[9] Die Antwort lautet, dass Armut Menschen unglücklich macht und dass Reichtum zwar die Lebenszufriedenheit steigern mag, aber das erlebte Wohlbefinden (im Schnitt) nicht verbessert.

Große Armut verstärkt die psychische Wirkung anderer Missgeschicke im Leben. Insbesondere Krankheiten sind für Menschen, die in großer Armut leben, viel schlimmer als für diejenigen, denen es besser geht.[10] Kopfschmerzen erhöhen bei Personen in den oberen zwei Dritteln der Einkommensverteilung den Anteil derjenigen, die über Trauer und Sorge berichten, von 19 auf 38 Prozent. Die entsprechenden Zahlen für das ärmste Zehntel sind 38 und 70 Prozent – ein höheres Ausgangsniveau und eine viel größere Steigerung. Auch bei den Folgen von Scheidung und

Einsamkeit sind erhebliche Unterschiede zwischen den Allerärmsten und den Übrigen festzustellen. Zudem sind die günstigen Auswirkungen des Wochenendes auf das erlebte Wohlergehen bei den Ärmsten deutlich geringer als für alle anderen.

Das Sättigungsniveau, ab dem das erlebte Wohlbefinden nicht weiter ansteigt, liegt bei einem Haushaltseinkommen von etwa 75 000 Dollar in Gebieten mit hohen Lebenshaltungskosten (in Gebieten, in denen die Lebenshaltungskosten niedriger sind, könnte es darunterliegen).[11] Die durchschnittliche Zunahme des erlebten Wohlbefindens, die mit Einkommen jenseits dieses Niveaus verbunden ist, betrug genau null. Dies ist erstaunlich, weil ein höheres Einkommen zweifellos den Kauf vieler Annehmlichkeiten, wie etwa Urlaubsaufenthalten an interessanten Orten oder Opernkarten, und verbesserte Lebensbedingungen ermöglicht. Weshalb schlagen sich diese zusätzlichen Annehmlichkeiten nicht in Berichten über die emotionale Erfahrung nieder? Nach einer plausiblen Interpretation geht ein höheres Einkommen mit einer verminderten Fähigkeit einher, die kleinen Freuden des Lebens zu genießen. Diese Annahme wird durch aussagekräftige Studienergebnisse erhärtet: Werden Studenten mit der Idee des Reichtums geprimt, verringert dies die Lust, die ihr Gesicht ausdrückt, wenn sie eine Tafel Schokolade essen![12]

Es besteht ein klarer Gegensatz zwischen den Effekten von Einkommen auf das erlebte Wohlbefinden und auf die Lebenszufriedenheit. Ein höheres Einkommen bringt eine höhere Zufriedenheit mit sich, weit über den Punkt hinaus, an dem es aufhört, irgendeinen positiven Effekt auf die Erfahrung zu haben. Die allgemeine Schlussfolgerung ist für das Wohlbefinden genauso klar wie für Darmspiegelungen: Die Bewertung der Lebensqualität und die tatsächliche Erfahrung mögen zusammenhängen, aber sie sind auch zwei verschiedene Dinge. Die Lebenszufriedenheit ist kein fehlerhaftes Maß des erlebten Wohlbefindens, wie ich vor ein paar Jahren glaubte. Sie ist etwas völlig anderes.

## Zum Thema »Erlebtes Wohlbefinden«

»Ziel der Politik sollte es sein, menschliches Leid zu verringern. Wir streben nach einem niedrigeren U-Index in der Gesellschaft. Die Bekämpfung von Depression und extremer Armut sollte eine Priorität sein.«

»Die Lebenszufriedenheit lässt sich am leichtesten dadurch steigern, dass man seine verfügbare Zeit besser nutzt. Können Sie mehr Zeit für Dinge erübrigen, die Ihnen Spaß machen?«

»Jenseits des Sättigungsniveaus des Einkommens können Sie sich zwar weitere angenehme Erfahrungen kaufen, aber Sie werden etwas von Ihrer Fähigkeit verlieren, die weniger teuren zu genießen.«

# *38. Lebenszufriedenheit*

Die Abbildung 16 ist einer Analyse von Andrew Clark, Ed Diener und Yannis Georgellis vom deutschen Sozio-ökonomischen Panel entnommen, im Zuge dessen dieselben Teilnehmer Jahr für Jahr nach ihrer Lebenszufriedenheit gefragt wurden.[1] Die Befragten berichteten auch über grundlegende Veränderungen in ihren Lebensumständen während des Vorjahres. Das Diagramm zeigt das Niveau der Zufriedenheit, das Menschen einige Jahre vor und nach ihrer Heirat angaben.

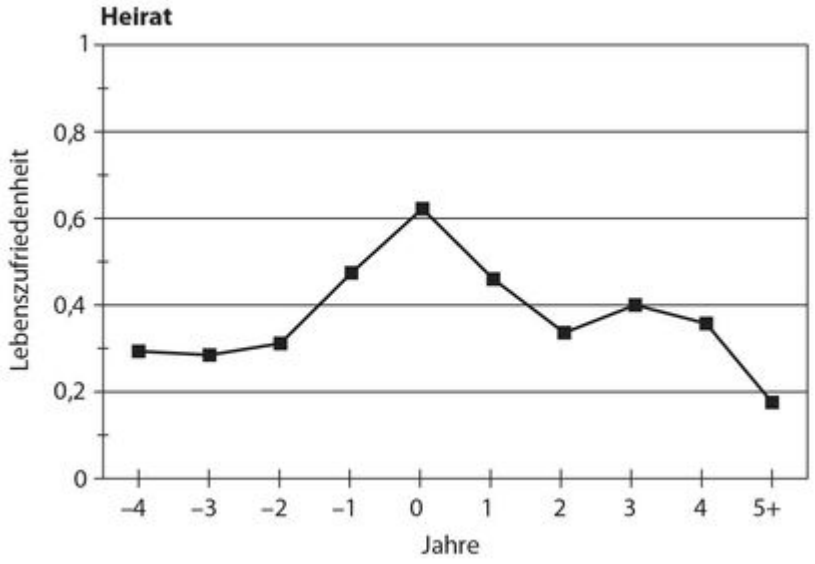


**Abbildung 16**

Das Diagramm sorgt immer wieder für nervöse Heiterkeit im Publikum, und die Nervosität ist leicht zu verstehen: Menschen, die beschließen zu heiraten, tun dies entweder in der Erwartung, sie würden dadurch glücklicher, oder weil sie hoffen, durch eine dauerhafte Bindung den gegenwärtigen Zustand der Seligkeit aufrechterhalten zu können. In dem Entschluss, zu heiraten, spiegelt sich bei vielen Menschen ein massiver affektiver Vorhersagefehler wider, um einen von Daniel Gilbert und Timothy Wilson geprägten nützlichen Begriff zu verwenden.[2] An ihrem Hochzeitstag wissen Braut und Bräutigam, dass die Scheidungsrate hoch

und die Häufigkeit der Enttäuschung in der Ehe sogar noch höher ist, aber sie glauben nicht, dass diese statistischen Befunde für sie gelten.

Die Erkenntnis, die Abbildung 16 veranschaulicht, ist der steile Rückgang der Lebenszufriedenheit. Das Diagramm wird für gewöhnlich so interpretiert, als bilde es einen Prozess der Anpassung ab, in dem die frühen Ehefreuden mit zunehmender Routine des Zusammenlebens schnell verblassen. Doch kann man die Daten auch anders interpretieren, indem man sich nämlich auf die Urteilsheuristik konzentriert. Wir fragen dann, was in den Köpfen der Menschen geschieht, wenn sie ihre Lebensqualität beurteilen sollen. Die Fragen »Wie zufrieden sind Sie mit Ihrem Leben insgesamt?« und »Wie glücklich sind Sie dieser Tage?« sind nicht so einfach zu beantworten wie »Wie lautet Ihre Telefonnummer?«. Wie gelingt es Umfrageteilnehmern, solche Fragen in wenigen Sekunden zu beantworten? Es ist hilfreich, sich auch diesen Vorgang als ein Urteil vorzustellen. Wie es auch bei anderen Fragen der Fall ist, mögen einige Menschen eine fertige Antwort darauf parat haben, die sie zu einer anderen Gelegenheit erzeugten, in der sie ihr Leben bewerteten. Andere, vermutlich die Mehrheit, finden nicht so schnell eine Antwort auf exakt die Frage, die ihnen gestellt wurde, und erleichtern sich die Aufgabe dadurch, dass sie eine andere Frage beantworten. Hier kommt System 1 ins Spiel. Wenn wir Abbildung 16 in diesem Licht betrachten, nimmt sie eine andere Bedeutung an.

Die Antworten auf viele einfache Fragen können eine globale Beurteilung der Lebensqualität ersetzen. Sie erinnern sich an die Studie, in der Studenten, die gerade gefragt worden waren, wie viele Verabredungen sie im letzten Monat hatten, ihre »gegenwärtige Zufriedenheit« so beurteilten, als wären Verabredungen die einzigen wichtigen Tatsachen in ihrem Leben.[3] In einem anderen bekannten Experiment in derselben Art baten Norbert Schwarz und seine Mitarbeiter Probanden in ihrem Forschungslabor, einen Fragebogen über ihre Lebenszufriedenheit auszufüllen.[4] Doch bevor sie sich dieser Aufgabe widmeten, sollten sie ein Blatt Papier für Schwarz fotokopieren. Die Hälfte der Befragten fand auf dem Kopierer ein Zehn-Cent-Stück, das von dem Experimentator dort hinterlegt worden war. Dieser unbedeutende glückliche Vorfall führte zu einer deutlichen Verbesserung der berichteten Zufriedenheit der Probanden

mit ihrem Leben insgesamt! Eine Stimmungsheuristik ist eine Methode, um Fragen nach der Lebenszufriedenheit zu beantworten.

Die Rendezvous-Umfrage und das Experiment mit der Münze auf dem Kopierer demonstrierten wie beabsichtigt, dass die Antworten auf Fragen nach dem globalen Wohlbefinden mit Vorsicht betrachtet werden sollten. Aber selbstverständlich ist Ihre gegenwärtige Stimmung nicht das Einzige, was Ihnen einfällt, wenn Sie gebeten werden, die Qualität Ihres Lebens zu beurteilen. Sie werden vermutlich an bedeutsame Ereignisse aus jüngster Vergangenheit oder naher Zukunft erinnert werden; an wiederkehrende Sorgen wie die Gesundheit Ihres Partners oder die schlechte Gesellschaft, in der sich Ihr Teenager bewegt; an wichtige Erfolge und schmerzliche Niederlagen. Ihnen werden einige Ideen durch den Kopf gehen, die für die Frage relevant sind; viele andere werden es nicht tun. Selbst wenn der Score, den wir unserer Lebenszufriedenheit zuschreiben, nicht von völlig belanglosen Zufällen wie der Münze auf dem Kopierer beeinflusst wird, wird er doch von einer kleinen Stichprobe hoch verfügbarer Ideen bestimmt und ist nicht das Resultat einer sorgfältigen Gewichtung unserer verschiedenen Lebensbereiche.

Menschen, die unlängst heirateten oder in naher Zukunft heiraten wollen, werden sich vermutlich an diese Tatsache erinnern, wenn ihnen eine allgemeine Frage über ihr Leben gestellt wird. Weil die Eheschließung in den Vereinigten Staaten fast immer aus freien Stücken erfolgt, wird sich praktisch jeder, der an seine kurz zurückliegende oder bevorstehende Heirat erinnert wird, über diesen Gedanken freuen. Aufmerksamkeit ist der Schlüssel für die Lösung des Rätsels. Abbildung 16 lässt sich als Diagramm der Wahrscheinlichkeit lesen, mit der Menschen an ihre kurz zurückliegende oder bevorstehende Heirat denken, wenn sie nach ihrer Lebenszufriedenheit gefragt werden. Die erhöhte Bewusstseinszugänglichkeit (Salienz) dieses Gedankens vermindert sich mit der Zeit, da seine Neuigkeit schwindet.

Die Abbildung zeigt ein ungewöhnlich hohes Niveau an Lebenszufriedenheit, das zwei bis drei Jahre um den Zeitpunkt der Eheschließung herum anhält. Doch wenn sich in diesem scheinbaren Anstieg der zeitliche Verlauf einer Heuristik für die Beantwortung der Frage widerspiegelt, können wir daraus nur wenig über das Lebensglück oder über den Prozess der Anpassung an die Ehe lernen. Wir können daraus nicht folgern, dass eine erhöhte Lebenszufriedenheit mehrere Jahre andauert und

dann nach und nach zurückgeht. Selbst Menschen, die mit Freude an ihre Ehe denken, wenn ihnen eine Frage nach ihrer Lebenszufriedenheit gestellt wird, sind normalerweise nicht unbedingt glücklicher. Wenn sie nicht einen Großteil des Tages an ihr Eheglück denken, wird dies ihr allgemeines Glücksgefühl nicht direkt beeinflussen. Selbst Frischverheiratete, die das Glück haben, dass ihre Liebe ihr Denken völlig beherrscht, werden eines Tages auf den Boden der Realität zurückkehren, und ihr erlebtes Wohlbefinden wird, wie beim Rest von uns, von dem Umfeld und den Aktivitäten des gegenwärtigen Augenblicks abhängen.

In den DRM-Studien gab es im Großen und Ganzen keinen Unterschied im erlebten Wohlbefinden zwischen Frauen mit Partner und Frauen ohne Partner. Dieses Ergebnis lässt sich mit der Art und Weise erklären, wie die beiden Gruppen ihre Freizeit nutzen. Frauen mit Partnern verbringen weniger Zeit allein, aber auch weniger Zeit mit Freunden. Sie verbringen mehr Zeit mit Sex, was wunderbar ist, aber sie wenden auch mehr Zeit für Hausarbeit, Kochen und Kinderbetreuung auf, bei denen es sich um eher unbeliebte Aktivitäten handelt. Und natürlich ist die viele Zeit, die verheiratete Frauen mit ihrem Ehemann verbringen, für einige viel angenehmer als für andere. Eine Eheschließung lässt das erlebte Wohlbefinden normalerweise unbeeinflusst, nicht weil die Heirat sich nicht auf die Zufriedenheit auswirken würde, sondern weil sie einige Aspekte des Lebens zum Besseren und andere zum Schlechteren verändert.

Eine Ursache für die niedrigen Korrelationen zwischen den individuellen Lebensverhältnissen und der Lebenszufriedenheit ist die Tatsache, dass sowohl Glückserleben als auch Lebenszufriedenheit weitgehend von dem genetisch verankerten Temperament bestimmt werden. Eine Disposition für psychisches Wohlergehen ist genauso erblich wie die Körpergröße oder die Intelligenz, wie Studien über Zwillinge zeigen, die getrennt aufwachsen. Menschen, die gleich wohlhabend sind, unterscheiden sich sehr stark in ihrem Glücksgefühl. In einigen Fällen, etwa bei Heirat, sind die Korrelationen mit dem Wohlbefinden aufgrund von Ausgleichseffekten niedrig. Die gleiche Situation mag für einige Menschen gut und für andere schlecht sein, und neue Lebensumstände haben sowohl Vorteile als auch Kosten. In anderen Fällen, etwa bei hohem Einkommen, sind die Auswirkungen auf die Lebenszufriedenheit im Allgemeinen positiv, aber das Bild wird durch die Tatsache kompliziert, dass manchen Menschen Geld viel mehr bedeutet als anderen.

Eine groß angelegte Studie über die Auswirkungen einer Universitätsausbildung, die zu einem anderen Zweck durchgeführt wurde, erbrachte bemerkenswerte Belege für die lebenslangen Auswirkungen der Ziele, die sich junge Menschen setzen.[5] Die entsprechenden Daten wurden Fragebogen entnommen, die von 1995 bis 1997 an etwa 12 000 Personen ausgegeben wurden, die 1976 ihr Studium an einer Eliteuniversität aufgenommen hatten. Mit 17 oder 18 Jahren hatten die Teilnehmer einen Fragebogen ausgefüllt, in den sie das Ziel, »einmal sehr wohlhabend zu sein«, auf einer Vier-Punkte-Skala einordnen sollten, die von »unwichtig« bis »sehr wichtig« reichte.[6] Der Fragebogen, den sie zwanzig Jahre später ausfüllten, erfasste ihr Einkommen im Jahr 1995 sowie ihre globale Lebenszufriedenheit.

Ziele machen einen großen Unterschied. 19 Jahre nachdem sie ihre finanziellen Ambitionen zu Protokoll gegeben hatten, hatten viele derjenigen, die ein hohes Einkommen anstrebten, dieses tatsächlich erreicht. Bei den 597 Ärzten und anderen medizinischen Fachkräften in der Stichprobe zum Beispiel war jeder zusätzliche Punkt auf der Geldwichtigkeits-Skala mit einer Zunahme des Arbeitseinkommens von über 14 000 Dollar (Wert bezogen auf das Jahr 1995) verbunden! Nicht erwerbstätige verheiratete Frauen hatten ihre finanziellen Ambitionen ebenfalls mit hoher Wahrscheinlichkeit befriedigt. Jeder Punkt auf der Skala entsprach über 12 000 Dollar zusätzlichem Haushaltseinkommen für diese Frauen, ganz offensichtlich dank des Verdienstes ihrer Ehemänner.

Die Bedeutung, die 18-Jährige dem Einkommen zuschrieben, nahm auch ihre Zufriedenheit mit ihrem Einkommen als Erwachsene vorweg. Wir verglichen die Lebenszufriedenheit in einer Gruppe mit hohem Einkommen (über 200 000 Dollar Haushaltseinkommen) mit der in einer Gruppe mit niedrigem bis mittlerem Einkommen (weniger als 50 000 Dollar). Der Effekt des Einkommens auf die Lebenszufriedenheit war bei denjenigen größer, die ein großes Vermögen als ein wichtiges Lebensziel angegeben hatten: 0,57 auf einer Fünf-Punkte-Skala. Die entsprechende Differenz für diejenigen, die angegeben hatten, Geld sei für sie nicht wichtig, betrug nur 0,12.[7] Menschen, die ein hohes Einkommen anstrebten und es auch hatten, waren erheblich zufriedener als der Durchschnitt; diejenigen, die ein hohes Einkommen anstrebten, aber nicht hatten, waren erheblich unzufriedener. Der gleiche Grundsatz gilt für andere Ziele – ein Rezept für ein

unglückliches Erwachsenenalter besteht darin, Ziele festzusetzen, die besonders schwer zu erreichen sind. Gemessen an der Lebenszufriedenheit zwanzig Jahre später war das am wenigsten erfolgversprechende Ziel, das sich ein junger Mensch setzen konnte, »es in einer darstellenden Kunst zur Meisterschaft zu bringen«. Die Ziele von Teenagern beeinflussen das, was ihnen widerfährt, wohin sie es bringen und wie zufrieden sie sind.

Nicht zuletzt aufgrund dieser Ergebnisse habe ich meine Meinung über die Definition von seelischem Wohlbefinden geändert. Die Ziele, die sich Menschen für sich selbst setzen, sind so wichtig für das, was sie tun, und wie sie es gefühlsmäßig erleben, dass die ausschließliche Konzentration auf das erlebte Wohlbefinden nicht vertretbar ist. Ein Konzept des Wohlbefindens, das die Wünsche von Menschen ignoriert, ist unhaltbar. Andererseits ist ein Konzept des Wohlbefindens, das die gelebte Erfahrung ignoriert und sich nur darauf konzentriert, wie Menschen ihre Stimmungslage beschreiben, wenn sie darüber nachdenken, genauso unhaltbar. Wir müssen uns mit der Komplexität einer hybriden Sichtweise abfinden, bei der das Wohlbefinden beider Selbste berücksichtigt wird.

## Die Fokussierungs-Illusion

Aus der Schnelligkeit, mit der Menschen auf Fragen nach ihrer Lebenszufriedenheit antworten, und aus dem Einfluss der aktuellen Stimmungslage auf ihre Antworten können wir folgern, dass sie bei der Beurteilung ihrer Lebenszufriedenheit keine sorgfältige Prüfung vornehmen. Sie benutzen Heuristiken, die Beispiele für Ersetzung und Berücksichtigung der aktuell verfügbaren Information (WYSIATI-Regel) sind. Obgleich ihre Einschätzung ihrer Lebenslage von einer Frage nach Verabredungen oder von einer Münze auf einem Kopierer beeinflusst wurde, vergaßen die Teilnehmer dieser Studien nicht, dass das Leben sich nicht auf Verabredungen oder glückliche Zufälle beschränkt. Das Glückskonzept ändert sich durch das Finden eines Zehn-Cent-Stücks nicht plötzlich, aber System 1 ersetzt bereitwillig das Ganze durch einen kleinen Teil. Jeder Aspekt des Lebens, auf den die Aufmerksamkeit gerichtet wird, spielt bei einer globalen Bewertung eine große Rolle. Dies ist das Wesen der Fokussierungs-Illusion, die sich in einem Satz zusammenfassen lässt:

> Nichts im Leben ist so wichtig, wie man glaubt, wenn man darüber nachdenkt.

Der Ursprung dieser Idee war eine familiäre Diskussion über den Umzug von Kalifornien nach Princeton; dabei behauptete meine Frau, in Kalifornien seien die Menschen glücklicher als an der Ostküste. Ich behauptete, das Klima sei nachweislich keine wichtige Bestimmungsgröße des Wohlbefindens – die Menschen in den skandinavischen Ländern seien wahrscheinlich die glücklichsten Menschen auf der ganzen Welt. Ich wies darauf hin, dass dauerhafte Lebensumstände keinen nennenswerten Einfluss auf das Wohlbefinden hätten, und versuchte meine Frau vergeblich davon zu überzeugen, dass ihre Intuitionen über das glückliche Leben der Kalifornier auf einem affektiven Vorhersagefehler beruhten.[8]

Kurze Zeit später, als mir diese Diskussion noch in frischer Erinnerung war, nahm ich an einem Workshop über die sozialwissenschaftlichen Dimensionen der globalen Erwärmung teil. Ein Kollege brachte ein Argument vor, das auf seiner Sichtweise des Wohlergehens der Weltbevölkerung in den nächsten hundert Jahren basierte. Ich entgegnete, es sei absurd, vorherzusagen, wie es sein wird, auf einem wärmeren Planeten zu leben, wenn wir nicht einmal wüssten, wie es ist, in Kalifornien zu leben. Kurz nach dieser Debatte wurden meinem Kollegen David Schkade und mir Forschungsgelder bewilligt, um zwei Fragen zu studieren: Sind Menschen, die in Kalifornien leben, glücklicher als andere? Und: Was sind die gängigen Ansichten über das relative Lebensglück der Kalifornier?

Wir sammelten umfangreiche Studentenstichproben an großen staatlichen Universitäten in Kalifornien, Ohio und Michigan. Von einigen erhielten wir einen detaillierten Bericht über ihre Zufriedenheit mit verschiedenen Aspekten ihres Lebens.[9] Von anderen bekamen wir eine Vorhersage darüber, wie jemand »mit Ihren Interessen und Wertvorstellungen«, der irgendwo anders lebt, den gleichen Fragebogen ausfüllen würde.

Als wir die Daten auswerteten, zeigte sich, dass ich in der familiären Kontroverse recht behalten hatte.[10] Wie erwartet, unterschieden sich die Studenten der beiden Regionen erheblich in ihrer Einstellung zu den jeweiligen klimatischen Bedingungen an ihren Wohnorten: Die Kalifornier genossen ihr Klima, die Mittelwestler verabscheuten das ihre. Aber das Klima war keine wichtige Determinante des Wohlbefindens. Tatsächlich bestand nicht der geringste Unterschied in der Lebenszufriedenheit von Studenten in Kalifornien und im Mittleren Westen.[11] Wir fanden auch heraus, dass meine Frau in ihrer Überzeugung, dass die Kalifornier glücklicher seien als andere, nicht allein war. Die Studenten in beiden Regionen teilten die gleiche irrige Auffassung, und wir konnten ihren Irrtum darauf zurückführen, dass sie dem Klima eine allzu hohe Bedeutung beimaßen. Wir nannten den Fehler »Fokussierungs-Illusion«.

Der Kern der Fokussierungs-Illusion ist die WYSIATI-Regel, die dazu führt, dass das Klima übergewichtet wird, während alle anderen Determinanten des Wohlbefindens zu niedrig gewichtet werden. Um zu ermessen, wie stark diese Illusion ist, nehmen Sie sich ein paar Sekunden Zeit, um diese Frage zu beantworten:

Wie viel Spaß macht Ihnen Ihr Auto?[12]

Ihnen fiel sofort eine Antwort ein; Sie wissen, wie sehr Sie Ihr Auto mögen und genießen. Betrachten Sie jetzt eine andere Frage: *»Wann* macht Ihnen Ihr Auto Spaß?« Die Antwort auf diese Frage mag Sie überraschen, aber sie ist eindeutig: Ihr Auto macht Ihnen Freude (oder Missvergnügen), wenn Sie über Ihr Auto nachdenken, was vermutlich nicht allzu oft geschieht. Unter normalen Umständen verbringen Sie nicht viel Zeit damit, über Ihr Auto nachzudenken, wenn Sie damit fahren. Sie denken beim Fahren an andere Dinge, und Ihre Stimmung ist abhängig davon, woran Sie denken. Wenn Sie versuchen würden, einzuschätzen, wie viel Spaß Ihnen Ihr Auto macht, würden Sie auch hier eine viel engere Frage beantworten: »Wie viel Spaß macht Ihnen Ihr Auto, *wenn Sie darüber nachdenken*?« Die Ersetzung veranlasste Sie dazu, die Tatsache zu ignorieren, dass Sie nur selten über Ihr Auto nachdenken, eine Art von Vernachlässigung der Dauer. Das Ergebnis ist eine Fokussierungs-Illusion. Wenn Sie Ihr Auto mögen, werden Sie wahrscheinlich das Vergnügen, das es Ihnen bereitet, zu hoch einschätzen, was Sie irreführen wird, wenn Sie über die Vorzüge Ihres gegenwärtigen Fahrzeugs nachdenken beziehungsweise wenn Sie erwägen, ein neues zu kaufen.

Eine ähnliche Verzerrung wirkt sich auf Urteile über die Lebenszufriedenheit der Kalifornier aus. Danach gefragt, beschwören Sie vermutlich ein Bild einer Person herauf, die sich einem typischen Aspekt des kalifornischen Lebensstils widmet, etwa Wandern im Sommer oder dem Genießen des milden Winterwetters. Die Fokussierungs-Illusion tritt auf, weil Kalifornier tatsächlich diesen Aspekten ihres Lebens wenig Aufmerksamkeit widmen. Zudem fällt langjährigen Kaliforniern vermutlich nicht das Klima ein, wenn sie um eine globale Beurteilung ihres Lebens gebeten werden. Wenn man sein ganzes Leben dort verbracht hat und nicht viel reist, ist in Kalifornien zu leben so, wie zehn Zehen zu haben: nett, aber nichts, worüber man viel nachdenkt. Gedanken über einen Aspekt des Lebens ziehen eher die Aufmerksamkeit auf sich, wenn eine kontrastierende Alternative in hohem Maße verfügbar ist.

Menschen, die sich erst vor Kurzem in Kalifornien niedergelassen haben, werden anders reagieren. Nehmen wir einen unternehmungslustigen Menschen, der aus Ohio zugezogen ist, weil er glaubt, dass er sich in einem besseren Klima wohler fühlen wird. Nach seinem Umzug wird ihn eine Frage nach seiner Lebenszufriedenheit vermutlich über etliche Jahre hinweg an den Wohnortswechsel erinnern und auch Gedanken an die gegensätzlichen klimatischen Bedingungen in den beiden Bundesstaaten wachrufen. Der Vergleich wird zweifellos Kalifornien begünstigen, und die Aufmerksamkeit auf diesen Aspekt des Lebens wird vielleicht dessen wahres Erfahrungsgewicht verzerren. Doch hat die Fokussierungs-Illusion auch eine angenehme Seite. Unabhängig davon, ob der Betreffende nach dem Umzug tatsächlich glücklicher ist, wird er sich selbst als glücklicher einschätzen, weil Gedanken an das Klima ihn zu der Überzeugung bringen werden, dass er es tatsächlich ist. Die Fokussierungs-Illusion kann dazu führen, dass sich Menschen über ihren gegenwärtigen Zustand des Wohlbefindens sowie über die Zufriedenheit anderer und über ihre eigene Zufriedenheit in der Zukunft täuschen.

> Welchen Prozentsatz des Tages verbringen Querschnittsgelähmte in einer schlechten Stimmung?[13]

Diese Frage löste in Ihnen höchstwahrscheinlich den Gedanken an einen Querschnittsgelähmten aus, der gegenwärtig über einen Aspekt seines Zustandes nachdenkt. Ihre Vermutung über die Stimmungslage eines Querschnittsgelähmten ist daher vermutlich für die erste Zeit nach dem schweren Unfall zutreffend; nach dem Ereignis denken Unfallopfer eine Zeit lang fast an nichts anderes. Doch von wenigen Ausnahmen abgesehen wird die Aufmerksamkeit mit der Zeit in dem Maße von der neuen Situation abgezogen, wie diese immer vertrauter wird. Die wichtigsten Ausnahmen stellen chronische Schmerzen, permanenter Lärm und schwere Depressionen dar. Schmerzen und Lärm sind biologisch verankerte Signale, die die Aufmerksamkeit auf sich ziehen, und eine Depression geht mit einem sich selbst verstärkenden Kreislauf negativer Gedanken einher. Daher gibt es keine Anpassung an diese Zustände. Eine

Querschnittslähmung ist jedoch keine dieser Ausnahmen: detaillierte Beobachtungen zeigen, dass sich Querschnittsgelähmte schon einen Monat nach ihrem Unfall mehr als die Hälfte der Zeit in einer recht guten Stimmung befinden – auch wenn sich ihre Stimmung mit Sicherheit stark eintrübt, sobald sie über ihre Situation nachdenken.[14] Die meiste Zeit aber arbeiten Querschnittsgelähmte oder sie lesen, amüsieren sich mit Freunden und werden wütend, wenn sie in der Zeitung Artikel über politische Inhalte lesen. Wenn sie irgendeiner dieser Aktivitäten nachgehen, unterscheiden sie sich kaum von allen anderen, und wir können erwarten, dass das erlebte Wohlbefinden von Querschnittsgelähmten die meiste Zeit in der Nähe des allgemeinen Durchschnitts liegt. Die Anpassung an eine positive oder negative neue Situation besteht zu einem großen Teil darin, immer weniger daran zu denken. In diesem Sinne sind die meisten langfristigen Lebensumstände wie Querschnittslähmung und Heirat Teilzeitzustände, in denen man sich immer nur dann befindet, wenn man die Aufmerksamkeit darauf richtet.

Eines der Privilegien eines Princeton-Professors ist die Gelegenheit, die Doktorarbeiten intelligenter Studenten zu betreuen. Und eines meiner Lieblingsexperimente in diesem Bereich war ein Projekt, bei dem Beruria Cohn Daten eines Meinungsforschungsinstituts, das die Befragten bat, den Prozentsatz der Zeit abzuschätzen, den Querschnittsgelähmte in schlechter Stimmung verbringen, zusammentrug und auswertete. Sie teilte die Befragten in zwei Gruppen ein: Einigen wurde gesagt, der schwere Unfall habe sich vor einem Monat ereignet, anderen wurde gesagt, er liege ein Jahr zurück. Zudem gab jeder Befragte an, ob er oder sie persönlich einen Querschnittsgelähmten kannte. Die Urteile der beiden Gruppen über diejenigen, die erst seit Kurzem querschnittsgelähmt waren, stimmten weitgehend überein: Diejenigen, die einen Querschnittsgelähmten kannten, schätzten, dass sie 75 Prozent der Zeit niedergeschlagen seien; diejenigen, die sich einen Querschnittsgelähmten vorstellen mussten, tippten auf 70 Prozent. Dagegen unterschieden sich die beiden Gruppen deutlich in ihren Einschätzungen der Stimmungslage von Querschnittsgelähmten ein Jahr nach den Unfällen: Diejenigen, die einen Querschnittsgelähmten kannten, schätzten, dass sie sich 41 Prozent der Zeit in gedrückter Stimmung befinden. Die Schätzungen derjenigen, die nicht persönlich mit einem Querschnittsgelähmten bekannt waren, beliefen sich auf durchschnittlich 68 Prozent. Offenbar hatten diejenigen, die einen Querschnittsgelähmten

kannten, beobachtet, dass dieser seine Aufmerksamkeit nach und nach von seiner Krankheit zurückzog. Urteile über die Stimmungslage von Lotteriegewinnern einen Monat und ein Jahr nach dem Ereignis zeigten genau das gleiche Muster.

Wir können davon ausgehen, dass die Lebenszufriedenheit von Querschnittsgelähmten und von Patienten mit anderen chronischen und belastenden Erkrankungen im Vergleich zu ihrem erlebten Wohlbefinden niedrig ist, weil die Aufforderung, ihre Lebensqualität zu beurteilen, sie notgedrungen an die Lebensqualität anderer und das Leben, das sie vor ihrer Erkrankung führten, erinnert. In Einklang mit dieser Annahme haben jüngste Studien über Kolostomie-Patienten (Patienten mit künstlichem Darmausgang) drastische Inkonsistenzen zwischen dem erlebten Wohlbefinden der Patienten und ihren Beurteilungen ihrer Lebensqualität zum Vorschein gebracht.[15] Erfahrungsstichproben zeigen keine Unterschiede im Glückserleben zwischen diesen Patienten und einer gesunden Population. Aber Kolostomie-Patienten wären bereit, für ein Leben ohne Kolostomie Lebensjahre herzugeben. Außerdem erinnern sich Patienten, deren Eingriff rückgängig gemacht wurde, voller Schrecken an die Zeit mit einem künstlichen Darmausgang, und sie würden sogar noch mehr von ihrer verbliebenen Lebenszeit opfern, um dies nicht noch einmal durchzumachen. Es hat hier den Anschein, als würde das erinnernde Selbst einer massiven Fokussierungs-Illusion über eine Lebensphase unterliegen, die das erlebende Selbst als ganz erträglich empfindet.

Daniel Gilbert und Timothy Wilson führten den Begriff »Misswollen« *(miswanting)* ein, um Fehlentscheidungen zu beschreiben, die das Ergebnis affektiver Vorhersagefehler sind.[16] Dieses Wort verdiente es, in die Alltagssprache Eingang zu finden. Die Fokussierungs-Illusion (die Gilbert und Wilson »Fokalismus« nennen) ist eine reiche Quelle von Misswollen. Insbesondere lässt sie uns die Bedeutung größerer Anschaffungen oder veränderter Lebensumstände für unser zukünftiges Wohlbefinden überschätzen.

Vergleichen Sie zwei Entscheidungen, die einige Aspekte Ihres Lebens verändern werden: der Kauf eines komfortablen neuen Autos und die Mitgliedschaft in einer Gruppe, die sich wöchentlich trifft, vielleicht einem Poker- oder Buchklub. Beide Erfahrungen werden am Anfang neu und spannend sein. Der entscheidende Unterschied besteht darin, dass Sie auf

längere Sicht dem Auto, wenn Sie es fahren, keine große Beachtung mehr schenken, während Sie Ihren sozialen Interaktionen immer Aufmerksamkeit schenken werden. Aufgrund der WYSIATI-Regel werden Sie wahrscheinlich den langfristigen Nutzen des Autos zu hoch einschätzen, während Sie für geselliges Beisammensein oder für an sich aufmerksamkeitsfordernde Aktivitäten wie Tennisspielen oder Cello-Unterricht nicht den gleichen Fehler machen werden. Die Fokussierungs-Illusion erzeugt eine kognitive Verzerrung zugunsten von Gütern und Erfahrungen, die zunächst lustvoll sind, aber mit der Zeit ihren Reiz verlieren. Die Zeit wird vernachlässigt, was dazu führt, dass Erfahrungen, die langfristig ihren Aufmerksamkeitswert behalten, nicht so geschätzt werden, wie sie es verdienten.

## Immer wieder der Faktor Zeit

In diesem Teil des Buches sind wir immer wieder auf die Rolle der Zeit gestoßen. Es ist logisch, die Zeiterfahrung des erlebenden Selbst als eine Folge von Momenten zu beschreiben, die jeweils einen bestimmten Wert besitzen. Der Wert einer Episode – ich habe sie Hedonimeter-Summe genannt – ist einfach die Summe der Werte ihrer Momente. Aber so werden die Episoden nicht mental repräsentiert. Das erinnernde Selbst, wie ich es beschrieben habe, erzählt ebenfalls Geschichten und trifft Entscheidungen, und weder die Geschichten noch die Entscheidungen repräsentierten die Zeit in der angemessenen Art und Weise. Im Erzählmodus wird eine Episode durch einige wenige entscheidende Momente repräsentiert, insbesondere Anfang, Höhepunkt und Schluss. Die Dauer wird vernachlässigt. Wir sahen diesen Fokus auf besondere Momente sowohl bei der Kalte-Hand-Situation als auch bei Violettas Geschichte.

Eine andere Form von Vernachlässigung der Dauer sahen wir in der Neuen Erwartungstheorie, in der ein Zustand durch den Übergang in diesen Zustand repräsentiert wird. Ein Lotteriegewinn erzeugt einen neuen Vermögenszustand, der eine Zeit lang anhalten wird, aber der Entscheidungsnutzen entspricht der vorweggenommenen Intensität der Reaktion auf die Neuigkeit, dass man gewonnen hat. Der Rückzug der Aufmerksamkeit und andere Anpassungen an den neuen Zustand werden vernachlässigt, da nur dieser kurze Zeitabschnitt betrachtet wird. Den gleichen Fokus auf den Übergang in den neuen Zustand und die gleiche Vernachlässigung von Zeit und Anpassung findet man bei Vorhersagen der Reaktion auf chronische Erkrankungen und natürlich bei der Fokussierungs-Illusion. Der Fehler, den Menschen bei der Fokussierungs-Illusion begehen, bezieht sich auf die Beachtung ausgewählter Momente und die Vernachlässigung dessen, was zu anderen Zeiten geschieht. Wir sind gute Geschichtenerzähler, aber unser Gehirn scheint nicht besonders gut für die Verarbeitung von Zeit ausgelegt zu sein.

Während der letzten zehn Jahre haben wir viele neue Erkenntnisse über das Glück gewonnen. Aber wir haben auch gelernt, dass das Wort »Glück« keine einfache Bedeutung hat und nicht so verwendet werden sollte, als

hätte es eine. Manchmal lässt uns der wissenschaftliche Fortschritt verwirrter zurück, als wir es vorher waren.

## Zum Thema »Lebenszufriedenheit«

»Sie dachte, es würde sie glücklich machen, sich ein modisches Auto zu kaufen, aber dies erwies sich als ein affektiver Vorhersagefehler.«

»Sein Auto hatte heute Morgen eine Panne, als er zur Arbeit fuhr, und er ist schlecht gelaunt. Dies ist kein guter Tag, um ihn nach seiner Arbeitszufriedenheit zu fragen!«

»Meistens wirkt sie recht fröhlich, aber wenn sie gefragt wird, sagt sie, sie wäre sehr unglücklich. Die Frage muss sie an ihre jüngste Scheidung erinnern.«

»Der Kauf eines größeren Hauses mag uns auf lange Sicht nicht zufriedener machen. Wir könnten an einer Fokussierungs-Illusion leiden.«

»Er hat beschlossen, seine Zeit zwischen zwei Städten aufzuteilen. Vermutlich ein schwerwiegender Fall von Misswollen.

# *Schlusswort*

Zu Beginn dieses Buches führte ich zwei fiktive Figuren ein, verbrachte einige Zeit damit, zwei Spezies zu diskutieren, und endete mit zwei Selbsten. Die beiden Figuren waren das intuitive System 1, das für das schnelle Denken zuständig ist, und das mehr Mühe erfordernde und gemächlichere System 2, das für das langsame Denken zuständig ist, System 1 überwacht und mit seinen beschränkten Ressourcen, so gut es geht, die Kontrolle aufrechterhält. Die beiden Spezies waren die fiktiven *Econs*, die im Land der Theorie lebten, und die *Humans*, die in der realen Welt handeln. Die beiden Selbste sind das erlebende Selbst, das die Erfahrungen macht, und das erinnernde Selbst, das Buch führt und die Entscheidungen trifft. In diesem Schlusskapitel betrachte ich einige Anwendungen dieser drei Unterscheidungen, wobei ich sie in umgekehrter Reihenfolge behandle.

## Zwei Selbste

Die Möglichkeit von Konflikten zwischen dem erinnernden Selbst und den Interessen des erlebenden Selbst erwiesen sich als ein schwierigeres Problem, als ich ursprünglich dachte. In einem frühen Experiment, der Kalte-Hand-Studie, führte die Kombination von Vernachlässigung der Dauer und Höchststand-Ende-Regel zu offensichtlich absurden Entscheidungen. Weshalb sollten sich Menschen freiwillig unnötigen Schmerzen aussetzen? Unsere Probanden überließen die Entscheidung dem erinnernden Selbst und zogen es vor, jenen Versuch zu wiederholen, der die bessere Erinnerung zurückließ, auch wenn er mit mehr Schmerzen verbunden war. Eine Entscheidung anhand der Qualität der Erinnerung zu treffen, mag in Extremfällen gerechtfertigt sein, zum Beispiel wenn posttraumatischer Stress droht, aber das Kalte-Hand-Experiment war nicht traumatisch. Ein objektiver Beobachter, der die Entscheidung für jemand anderen trifft, würde sich zweifellos für die kurze unangenehme Reizexposition entscheiden, also das erlebende Selbst des Leidenden begünstigen. Die Entscheidungen, die Menschen für sich selbst treffen, lassen sich durchaus zutreffend als Fehlentscheidungen bezeichnen. Die Vernachlässigung der Dauer und die Höchststand-Ende-Regel bei der Beurteilung von Geschichten sowohl in der Oper als auch bei Urteilen über Jens Leben sind ebenso unhaltbar. Es ist unsinnig, ein ganzes Leben nach seinen letzten Momenten zu beurteilen oder bei der Beantwortung der Frage, welches Leben wünschenswerter ist, der Dauer keine Bedeutung beizumessen.

Das erinnernde Selbst ist eine Konstruktion von System 2. Doch die spezifischen Merkmale der Art und Weise, wie es Episoden und Leben beurteilt, sind Eigenschaften unseres Gedächtnisses. Die Vernachlässigung der Dauer und die Höchststand-Ende-Regel haben ihren Ursprung in System 1 und entsprechen nicht unbedingt den Werten von System 2. Wir halten die Dauer für wichtig, aber unser Gedächtnis sagt uns, dass dies nicht der Fall ist. Die Regeln, nach denen die Vergangenheit beurteilt wird, sind eine schlechte Richtschnur für die Entscheidungsfindung, weil die Zeit eine Rolle spielt. Dass Zeit eine letztlich begrenzte Ressource ist, ist die zentrale Tatsache unseres Lebens, doch unser erinnerndes Selbst ignoriert diese

Tatsache. Die Vernachlässigung der Dauer verursacht in Verbindung mit der Höchststand-Ende-Regel eine Verzerrung, die eine kurze Phase intensiver Lust einer längeren Phase mäßigen Glücks vorzieht. Das Spiegelbild der gleichen Verzerrung lässt uns eine kurze Phase starker, aber aushaltbarer Schmerzen stärker fürchten als eine viel längere Phase mittlerer Schmerzen. Die Vernachlässigung der Dauer macht uns auch anfällig dafür, eine längere Phase leichter Unannehmlichkeit zu akzeptieren, weil das Ende besser sein wird, und sie begünstigt auch den Verzicht auf eine Gelegenheit für eine lange glückliche Phase, wenn diese mit hoher Wahrscheinlichkeit negativ enden wird. Um die gleiche Idee bis zur Unannehmlichkeit zu steigern, betrachten Sie die weitverbreitete Ermahnung: »Tun Sie das nicht, Sie werden es bereuen.« Der Rat hört sich klug an, weil antizipiertes Bedauern das Verdikt des erinnernden Selbst ist, und wir neigen dazu, solche Urteile als endgültig und abschließend hinzunehmen. Doch wir sollten nicht vergessen, dass die Perspektive des erinnernden Selbst nicht immer richtig ist. Ein objektiver Beobachter des Hedonimeter-Profils, der an die Interessen des erlebenden Selbst denkt, würde vielleicht durchaus einen anderen Rat geben. Die Vernachlässigung der Dauer des erinnernden Selbst, seine übertriebene Gewichtung von Höchstständen und Enden sowie seine Anfälligkeit für den Rückschaufehler führen insgesamt zu einer verzerrten Abbildung unserer tatsächlichen Erfahrungen.

Im Gegensatz dazu behandelt die nach Dauer gewichtete Konzeption des Wohlbefindens alle Momente des Lebens gleich, egal ob sie erinnernswert sind oder nicht. Einige Augenblicke werden schließlich stärker gewichtet als andere, entweder weil sie erinnernswert oder weil sie wichtig sind. Die Zeit, die Menschen in einem unvergesslichen Moment verweilen, sollte in seine Dauer einbezogen werden, was sein Gewicht erhöht. Ein Moment kann auch dadurch an Bedeutung gewinnen, dass er das Erleben nachfolgender Momente verändert. So kann zum Beispiel eine Stunde, in der man Geige übte, das Musikerleben viele Jahre später verändern – egal ob man selbst Musik macht oder nur Musik hört. In ähnlicher Weise sollte ein kurzes schreckliches Ereignis, das eine Posttraumatische Belastungsstörung verursacht, mit der Gesamtdauer des langfristigen psychischen Leids, das sie auslöst, gewichtet werden. In der nach Dauer gewichteten Sichtweise können wir erst im Nachhinein bestimmen, ob ein Moment unvergesslich oder bedeutungsvoll ist. Die Aussagen »Ich werde mich immer daran erinnern …« oder »Dies ist ein bedeutender Augenblick«

sollten als Versprechungen oder Vorhersagen betrachtet werden, die falsch sein können – und es oftmals sind –, selbst wenn sie völlig ernst gemeint sind. Es ist eine gute Wette, dass viele der Dinge, von denen wir sagen, dass wir uns immer daran erinnern werden, zehn Jahre später längst vergessen sein werden.

Die Logik der Gewichtung der Dauer ist zwingend, aber sie kann nicht als eine vollständige Theorie des Wohlbefindens betrachtet werden, weil Individuen sich mit ihrem erinnernden Selbst identifizieren und ihre Geschichte wichtig nehmen. Eine Theorie des Wohlbefindens, die die Wünsche von Menschen ignoriert, ist nicht haltbar. Andererseits ist eine Theorie, die ignoriert, was tatsächlich im Leben von Menschen passiert, und sich ausschließlich auf das konzentriert, was sie über ihr Leben denken, ebenfalls nicht haltbar. Sowohl das erinnernde Selbst als auch das erlebende Selbst müssen berücksichtigt werden, weil ihre Interessen nicht immer deckungsgleich sind. Philosophen könnten sich sehr lange mit diesen Fragen abmühen.

Die Frage, welches der beiden Selbste wichtiger ist, ist nicht nur eine Frage für Philosophen; sie hat Konsequenzen für mehrere Politikfelder, insbesondere für die Gesundheits- und die Sozialpolitik. Betrachten wir die Gelder, die für die Behandlung verschiedener Krankheiten wie Blindheit, Taubheit oder Nierenversagen ausgegeben werden sollten. Sollten die Investitionen davon abhängen, wie viele Menschen diese Erkrankungen fürchten? Sollten sich die Investitionen an dem tatsächlichen Leid der Patienten orientieren? Oder sollten sie der Stärke des Wunsches der Patienten folgen, Linderung von ihrem Leid zu erfahren, und der Opfer, die sie zu bringen bereit wären, um diese Linderung zu erreichen? Die Rangfolge von Blindheit und Taubheit oder von Kolostomie und Dialyse könnte unterschiedlich ausfallen, je nachdem, welches Maß für den Schweregrad des Leidens verwendet wird. Es ist keine leichte Lösung in Sicht, aber die Frage ist zu wichtig, als dass man sie ignorieren könnte.[1]

Die Möglichkeit, Maße des Wohlbefindens als Leitplanken für politische Entscheidungen heranzuziehen, hat in jüngster Zeit sowohl unter Akademikern als auch bei einigen europäischen Regierungen großes Interesse geweckt.[2] Es ist jetzt denkbar – was vor ein paar Jahren noch unmöglich erschien –, dass ein Index der Menge des Leidens in einer Gesellschaft eines Tages zusammen mit Kenngrößen der Arbeitslosigkeit,

der Körperbehinderung und des Volkseinkommens in die nationalen Statistiken aufgenommen werden wird. Dieses Projekt hat große Fortschritte gemacht.

## Econs und Humans

In der Alltagssprache bezeichnen wir Menschen dann als vernünftig *(reasonable)*, wenn man verständig mit ihnen reden kann, wenn ihre Überzeugungen im Allgemeinen mit der Realität in Einklang stehen und wenn ihre Präferenzen ihren Interessen und Werten entsprechen. Das Wort »rational« vermittelt eine Vorstellung von sorgfältigerer Überlegung, strengerem Kalkül und weniger Herzlichkeit, aber in der Alltagssprache gilt eine rationale Person zweifellos als vernünftig. Für Ökonomen und Entscheidungstheoretiker hat das Adjektiv eine völlig andere Bedeutung. Der einzige Rationalitätstest besteht nicht darin, ob die Überzeugungen und Präferenzen einer Person vernünftig sind, sondern darin, ob sie in sich widerspruchsfrei sind. Eine rationale Person kann an Geister glauben, solange alle ihre anderen Überzeugungen mit der Existenz von Geistern konsistent sind. Eine rationale Person kann es vorziehen, gehasst oder geliebt zu werden, solange ihre Präferenzen in sich widerspruchsfrei sind. Rationalität ist logische Kohärenz – ob vernünftig oder nicht. *Econs* sind definitionsgemäß rational, aber es gibt erdrückende Beweise dafür, dass *Humans* nicht rational sein können. Ein *Econ* wäre nicht anfällig für Priming, für die WYSIATI-Regel, enges Framing, die Innenperspektive oder eine Präferenzumkehr; Dinge, die *Humans* nicht konsequent vermeiden können.

Die Definition von Rationalität als Kohärenz ist in einer wirklichkeitsfremden Weise restriktiv; sie verlangt die Einhaltung von Regeln der Logik, die ein begrenzter Intellekt nicht leisten kann. Vernunftbegabte Menschen können nach dieser Definition nicht rational sein, aber sie sollten deshalb nicht als irrational gebrandmarkt werden. »Irrational« ist ein starkes Wort, in dem Impulsivität, Emotionalität und eine sture Unzugänglichkeit für wohlbegründete Argumente mitschwingen.[3] Ich zucke oft zusammen, wenn die wissenschaftlichen Beiträge, die ich zusammen mit Amos verfasste, mit den Worten gelobt werden, wir hätten dort den Nachweis erbracht, dass das menschliche Entscheidungsverhalten irrational sei, während unsere Forschungen in Wirklichkeit nur gezeigt haben, dass *Humans* sich mit dem Modell des rationalen Agenten nicht gut beschreiben lassen.

Auch wenn *Humans* nicht irrational sind, brauchen sie oft Hilfe, um zu treffenderen Urteilen und besseren Entscheidungen zu gelangen, und in manchen Fällen können politische Maßnahmen und Institutionen diese Hilfe bieten. Diese Behauptungen mögen als harmlos erscheinen, tatsächlich aber sind sie sehr umstritten. Der Glaube an die menschliche Rationalität, wie er von der bedeutenden Chicagoer Schule in den Wirtschaftswissenschaften vertreten wird, ist eng verbunden mit einer Ideologie, die es als unnötig und sogar unmoralisch betrachtet, Menschen vor ihren Entscheidungen zu schützen. Rationale Menschen sollten frei sein, und sie sollten in verantwortungsvoller Weise ihre Angelegenheiten selbst regeln können. Milton Friedman, der führende Vertreter dieser Schule, drückte seine Sichtweise im Titel eines seiner populärsten Bücher aus: *Free to Choose* (dt. *Chancen, die ich meine*).

Die Annahme, dass Menschen rational handeln, bildet die intellektuelle Grundlage der libertären politischen Philosophie: Die individuelle Entscheidungsfreiheit soll nicht eingeschränkt werden, solange die Entscheidungen des Einzelnen andere nicht schädigen. Weiteren Rückhalt findet die libertäre Politik in der Bewunderung für die Effizienz von Märkten bei der Zuteilung von Gütern an die Konsumenten, die bereit sind, die höchsten Preise dafür zu zahlen. Ein berühmtes Beispiel für die Chicagoer Betrachtungsweise ist ein Beitrag mit dem Titel »A Theory of Rational Addiction« (»Eine Theorie rationalen Suchtverhaltens«); darin wird erklärt, wie ein rationaler Agent mit einer ausgeprägten Präferenz für starke und unmittelbare Gratifikationen die rationale Entscheidung treffen mag, zukünftige Sucht als eine Konsequenz hinzunehmen. [4] Gary Becker, einer der Verfasser dieses Beitrags, ein Nobelpreisträger für Wirtschaftswissenschaften und Anhänger der Chicagoer Schule, hörte ich einmal mit einem leichten Augenzwinkern, aber nicht gänzlich scherzhaft behaupten, wir sollten die Möglichkeit erwägen, die sogenannte Fettleibigkeitsepidemie mit dem Glauben von Menschen zu erklären, dass Diabetes schon bald heilbar sein wird. Er brachte ein triftiges Argument: Wenn wir Menschen beobachten, die sich scheinbar seltsam verhalten, sollten wir zunächst die Möglichkeit prüfen, dass sie einen guten Grund dafür haben, sich so zu verhalten. Psychologische Interpretationen sollten nur ins Feld geführt werden, wenn die Gründe unglaubwürdig werden – was Beckers Erklärung der Fettleibigkeit vermutlich ist.

In einer Nation von *Econs* sollte sich der Staat heraushalten und den *Econs* erlauben, zu tun und zu lassen, was ihnen gefällt, solange sie dadurch anderen keinen Schaden zufügen. Wenn ein Motorradfahrer beschließt, ohne Helm zu fahren, wird ein Anhänger der libertären Anschauung für sein Recht eintreten, dies zu tun. Bürger wissen, was sie tun, selbst wenn sie sich entscheiden, nicht fürs Alter vorzusorgen, oder wenn sie suchterzeugende Substanzen zu sich nehmen. Diese Position hat manchmal unangenehme Folgen: Ältere Menschen, die keine ausreichende Altersvorsorge betrieben haben, erhalten wenig mehr Mitgefühl als jemand, der sich nach dem Verzehr eines üppigen Mahls in einem Restaurant über die Rechnung beschwert. In der Debatte zwischen der Chicagoer Schule und den Verhaltensökonomen, die die extreme Form des Modells vom rationalen Agenten ablehnen, steht daher viel auf dem Spiel. Freiheit ist dabei kein strittiger Wert; alle an der Debatte Teilnehmenden sind dafür. Aber für Verhaltensökonomen ist das Leben komplexer als für diejenigen, die uneingeschränkt an die Rationalität des Menschen glauben. Kein Verhaltensökonom befürwortet einen Staat, der seine Bürger dazu zwingt, sich ausgewogen zu ernähren und sich nur erbauliche Fernsehsendungen anzusehen. Aber für Verhaltensökonomen hat die Freiheit Kosten. Diese werden von Personen getragen, die schlechte Entscheidungen treffen, und von einer Gesellschaft, die sich gezwungen fühlt, ihnen zu helfen. Die Frage, ob man Personen vor Fehlentscheidungen schützen sollte oder nicht, stellt daher für Verhaltensökonomen ein Dilemma dar. Die Ökonomen der Chicagoer Schule kennen dieses Problem nicht, weil rationale Agenten keine Fehler machen. Für Anhänger dieser Schule ist die Freiheit kostenlos.

Im Jahr 2008 taten sich der Volkswirt Richard Thaler und der Jurist Cass Sunstein zusammen, um ein Buch mit dem Titel *Nudge* zu schreiben, das bald zu einem internationalen Bestseller und zur Bibel der Verhaltensökonomik wurde.[5] Ihr Buch führte mehrere neue Wörter in die Alltagssprache ein, darunter auch *Econs* und *Humans*. Es präsentierte auch eine Reihe von Lösungen für das Dilemma, wie man Menschen helfen kann, gute Entscheidungen zu treffen, ohne ihre Freiheit einzuschränken. Thaler und Sunstein vertreten eine Position des libertären Paternalismus, wonach es dem Staat und anderen Institutionen erlaubt ist, Bürger zu solchen Entscheidungen *anzustoßen*, die ihren eigenen langfristigen Interessen dienlich sind. Ein Beispiel für einen solchen Anstoß ist es, wenn die Teilnahme an einer zusätzlichen privaten Altersvorsorge als

Standardoption festgelegt wird. Es lässt sich nur schwer behaupten, dass die persönliche Freiheit dadurch geschmälert wird, dass man automatisch in einen Plan zur Altersvorsorge aufgenommen wird, wenn man nur ein Kästchen ankreuzen muss, um nicht daran teilzunehmen. Wie wir früher sahen, wirkt sich das Framing der individuellen Entscheidung – Thaler und Sunstein nennen es »Entscheidungsarchitektur« – in erheblichem Maße auf das Ergebnis aus. Der Entscheidungsanstoß basiert auf soliden psychologischen Grundlagen, die ich weiter vorn beschrieben habe. Die Standardoption wird selbstverständlich als die normale Entscheidung betrachtet. Die Abweichung von der normalen Alternative ist ein aktives Handeln, das anstrengenderes Nachdenken erfordert, mit mehr Verantwortung verbunden ist und eher Reue hervorrufen wird, als nichts zu tun. Dies sind mächtige Kräfte, die die Entscheidung von jemandem anleiten können, der ansonsten unsicher ist, was er tun soll.

Mehr als *Econs* brauchen *Humans* auch Schutz vor anderen, die gezielt ihre Schwächen ausnutzen – insbesondere die Eigenarten von System 1 und die Trägheit von System 2. Bei rationalen Agenten wird davon ausgegangen, dass sie wichtige Entscheidungen nach reiflicher Überlegung treffen und alle Informationen nutzen, die ihnen zur Verfügung gestellt werden. Ein *Econ* liest und versteht das Kleingedruckte eines Vertrages, eher er diesen unterzeichnet, während *Humans* dies für gewöhnlich nicht tun. Eine skrupellose Firma, die Verträge entwirft, welche Kunden ohne gründliches Durchlesen routinemäßig unterzeichnen, hat erheblichen rechtlichen Gestaltungsspielraum, wichtige Informationen vor aller Augen zu verbergen. Eine schädliche Folge des Modells vom rationalen Agenten in seiner Extremform ist die Annahme, Kunden bräuchten keinen Schutz, sofern sichergestellt sei, dass alle relevanten Informationen offengelegt würden. Die Größe der Schrift und die Komplexität der Sprache bei der Offenlegung gelten nicht als relevant – ein *Econ* weiß, wie er mit dem Kleingedruckten klarkommt, wenn es von Belang ist. Dagegen verlangen die Empfehlungen in dem Buch *Nudge* von Firmen, Verträge anzubieten, die von *Humans* einigermaßen leicht zu lesen und zu verstehen sind. Es ist ein gutes Zeichen, dass einige dieser Empfehlungen auf den entschiedenen Widerstand von Firmen gestoßen sind, deren Gewinne vielleicht geschmälert würden, wenn ihre Kunden besser informiert wären. Eine Welt, in der Firmen miteinander konkurrieren, indem sie bessere Produkte

anbieten, ist einer Welt vorzuziehen, in der jene Firma die höchsten Gewinne macht, die sich am besten auf Verschleierung versteht.

Ein bemerkenswertes Merkmal des libertären Paternalismus ist die Anziehungskraft, die er auf ein breites politisches Spektrum ausübt. Das Paradebeispiel einer verhaltensökonomischen gesetzlichen Maßnahme, der sogenannte »Save More Tomorrow«-Plan, wurde im US-Kongress von einer ungewöhnlichen Koalition getragen, der extreme Konservative genauso angehörten wie Linke. Der »Save More Tomorrow«-Plan dient dem Vermögensaufbau, und Firmen können ihn ihren Mitarbeitern anbieten. Diejenigen, die daran teilnehmen, erlauben ihrem Arbeitgeber, jedes Mal, wenn sie eine Lohn- oder Gehaltserhöhung bekommen, ihren Beitrag zu ihrem Sparplan um einen festen Prozentsatz zu erhöhen. Die Erhöhung der Sparquote erfolgt automatisch, bis der Beschäftigte mitteilt, dass er in Zukunft darauf verzichten will. Diese brillante Innovation, die 2003 von Richard Thaler und Shlomo Benartzi vorgeschlagen wurde, hat mittlerweile die Sparquote und die zukünftige finanzielle Absicherung von Millionen von Arbeitnehmern verbessert. Sie basiert auf den psychologischen Prinzipien, die den Lesern dieses Buches mittlerweile vertraut sind. Sie vermeidet den Widerstand gegen einen unmittelbaren Verlust, indem sie keine unmittelbare Veränderung erfordert; dadurch, dass sie eine Erhöhung der Sparleistung an Lohnsteigerungen knüpft, verwandelt sie Verluste in entgangene Gewinne, die viel leichter zu ertragen sind; und die Automatik bringt die Trägheit von System 2 mit den langfristigen Interessen der Arbeitnehmer zur Deckung. Dabei wird selbstverständlich niemand zu irgendetwas gezwungen, was er nicht will, und niemand wird irregeführt oder manipuliert.

Die Anziehungskraft des libertären Paternalismus wurde in vielen Ländern – unter anderem Großbritannien und Südkorea – und von Politikern unterschiedlichster Couleur, wie etwa den britischen Konservativen und der von den Demokraten getragenen Regierung von Präsident Obama, erkannt. Tatsächlich hat die britische Regierung eine kleine Dienststelle eingerichtet, deren Aufgabe es ist, verhaltenswissenschaftliche Prinzipien anzuwenden, um der Regierung zu helfen, ihre Ziele besser zu erreichen. Der offizielle Name dieser Gruppe lautet »Behavioural Insight Team«, aber innerhalb und außerhalb der Regierung firmiert sie einfach als *Nudge Unit*. Thaler ist ein Berater dieses Teams.

In einer bilderbuchhaften Fortsetzung zu *Nudge* erhielt Sunstein von Präsident Obama das Angebot, die Leitung der US-Behörde für Informationspolitik und Regulierung zu übernehmen – eine Position, die ihm beträchtliche Möglichkeiten gab, die Anwendung von Erkenntnissen der Psychologie und Verhaltensökonomik in Behörden zu fördern. Diese Aufgabe wird im Rechenschaftsbericht für das Jahr 2010 des Office of Management and Budget – der obersten US-Bundesbehörde für Verwaltungs- und Budgetaufsicht – beschrieben. Leser dieses Buches werden die Logik hinter spezifischen Empfehlungen verstehen, wie etwa der Förderung einer »klaren, einfachen, einprägsamen und bedeutungsvollen öffentlichen Berichterstattung [über die Regierungsarbeit]«. Auch Hintergrundaussagen werden ihnen bekannt vorkommen, wie etwa: »Die Präsentation ist von großer Bedeutung; wird beispielsweise ein potenzielles Ergebnis als ein Verlust dargestellt, so wird sich dies unter Umständen stärker auswirken, als wenn es als ein Gewinn präsentiert wird.«

Das Beispiel einer Regulierung über das Framing von Angaben zum Kraftstoffverbrauch wurde bereits erwähnt. Zu weiteren Anwendungen, die realisiert wurden, gehört die Krankenversicherungspflicht, eine neue Version von Ernährungsrichtlinien, die die unverständliche Ernährungspyramide durch das anschauliche Bild eines Nahrungstellers mit ausgewogener Kost ersetzt, und eine vom US-Landwirtschaftsministerium aufgestellte Vorschrift, die es erlaubt, auf dem Etikett von Fleischprodukten Angaben wie etwa »90 Prozent fettfrei« zu machen, vorausgesetzt die Angabe »10 Prozent Fett« wird »direkt daneben, in einer Beschriftung der gleichen Farbe, Größe und Schriftart und auf dem gleichen farblichen Hintergrund wie die Angabe der Fettfreiheit in Prozent gezeigt«. Anders als *Econs* brauchen *Humans* Hilfe, um gute Entscheidungen zu treffen, und es gibt sachkundige und unaufdringliche Möglichkeiten, dies zu tun.

## Zwei Systeme

In diesem Buch wurde die Funktionsweise des menschlichen Intellekts *(mind)* als eine gespannte Interaktion zwischen zwei fiktiven Figuren beschrieben: dem automatischen System 1 und dem »anstrengenden« – eine gezielte willentliche Aktivierung erfordernden – System 2. Sie sind mittlerweile recht gut vertraut mit den Persönlichkeiten der beiden Systeme und können vorhersehen, wie sie wahrscheinlich in verschiedenen Situationen reagieren werden. Und selbstverständlich erinnern Sie sich auch daran, dass die beiden Systeme im Gehirn oder irgendwo anders nicht real existieren. »System 1 tut X« ist eine Abkürzung für »X ereignet sich automatisch«. Und »System 2 wird mobilisiert, um Y zu tun« ist eine Abkürzung für »Der allgemeine Aktivierungsgrad des zentralen Nervensystems *(arousal)* steigt, die Pupillen erweitern sich, die Aufmerksamkeit ist fokussiert, und Aktivität Y wird ausgeführt«. Ich hoffe, Sie finden die System-Metapher genauso hilfreich wie ich und Sie haben eine intuitive Vorstellung darüber entwickelt, wie die Systeme funktionieren, ohne sich durch die Frage verwirren zu lassen, ob sie existieren. Nachdem ich diese notwendige Warnung ausgesprochen habe, werde ich diese Terminologie bis zum Schluss beibehalten.

Das aufmerksamkeitsgesteuerte System 2 ist das, was wir als unser bewusstes Selbst betrachten. System 2 äußert Urteile und trifft Entscheidungen, aber es unterstützt oder rationalisiert oftmals Vorstellungen und Gefühle, die von System 1 erzeugt wurden. Sie mögen nicht wissen, dass Sie ein Projekt optimistisch einschätzen, weil etwas an dessen Leiterin Sie an Ihre geliebte Schwester erinnert, oder dass Sie eine Person unsympathisch finden, die entfernt Ihrem Zahnarzt ähnlich sieht. Doch wenn man Sie nach einer Erklärung fragt, werden Sie Ihr Gedächtnis nach plausiblen Gründen durchforsten und mit Sicherheit einige finden. Außerdem werden Sie die Geschichte glauben, die Sie erfunden haben. Aber System 2 ist nicht nur ein Verteidiger von System 1; es verhindert auch, dass viele verrückte Gedanken und nicht situationsadäquate Impulse offen zum Ausdruck gebracht werden. Die Investition von Aufmerksamkeit verbessert die Leistung bei zahlreichen Aktivitäten – denken Sie an die Risiken des Fahrens durch eine verengte Stelle, wenn Sie zerstreut sind –

und ist bei einigen Aufgaben wie etwa Vergleichen, Entscheiden und geordnetem Denken unverzichtbar. Aber System 2 ist kein Inbegriff von Rationalität. Seine Fähigkeiten sind begrenzt, und das Gleiche gilt für das Wissen, zu dem es Zugang hat. Wir denken nicht immer streng logisch, wenn wir nachdenken, und die Fehler sind nicht immer auf falsche Intuitionen zurückzuführen, die sich von selbst aufdrängen. Oft machen wir Fehler, weil wir (unser System 2) es nicht besser wissen.

Ich habe der Beschreibung von System 1 breiten Raum gegeben, und ich habe intuitiven Urteils- und Entscheidungsfehlern, die ich auf System 1 zurückführe, viele Seiten gewidmet. Aber die relative Anzahl von Seiten ist ein schlechter Indikator für die Bilanz zwischen den Glanzleistungen und den Fehlern des intuitiven Denkens. Tatsächlich hat ein Großteil dessen, was wir falsch machen, seinen Ursprung in System 1, aber System 1 ist auch der Ursprung der meisten Dinge, die wir richtig machen – und das ist das meiste dessen, was wir tun. Unsere Gedanken und Handlungen werden routinemäßig von System 1 gesteuert, und sie liegen im Allgemeinen richtig. Eine der Glanzleistungen ist das differenzierte und detaillierte Modell unserer Welt, das im assoziativen Gedächtnis abgespeichert ist: Es unterscheidet im Bruchteil einer Sekunde überraschende von normalen Ereignissen, erzeugt sofort eine Vorstellung von dem, was anstatt einer Überraschung erwartet wurde, und sucht automatisch nach einer kausalen Interpretation von Überraschungen und Ereignissen, sowie sie stattfinden.

Im Gedächtnis ist auch das gewaltige Repertoire an Fähigkeiten gespeichert, die wir in lebenslanger Übung erworben haben und die automatisch geeignete Lösungen für unvermittelt auftretende Herausforderungen generieren, vom Umgehen eines großen Steins, der im Weg liegt, bis zur Abwendung des beginnenden Wutanfalls eines Kunden. Der Erwerb von Fähigkeiten erfordert ein geregeltes Umfeld, ausreichende Übungsgelegenheiten und zügige sowie unzweideutige Rückmeldungen über die Richtigkeit von Gedanken und Handlungen. Wenn diese Bedingungen erfüllt sind, entwickelt sich langfristig eine Fähigkeit, und die intuitiven Urteile und Entscheidungen, die uns dann spontan einfallen, werden überwiegend zutreffend sein. All dies leistet System 1, was bedeutet, dass es sich um eine automatische und zügige Aktivität handelt. Ein Kennzeichen von sachkundigem Leistungsvermögen ist die Fähigkeit, große Mengen an Informationen schnell und effizient zu verarbeiten.

Wenn man einer Herausforderung begegnet, für die eine sachkundige Reaktion verfügbar ist, wird diese Reaktion ausgelöst. Was geschieht bei fehlender Sachkunde? Manchmal, etwa bei dem Problem 17 × 24 = ?, das eine spezifische Antwort verlangt, ist es unmittelbar offensichtlich, dass System 2 einbezogen werden muss. Aber System 1 ist nur selten ratlos; es wird nicht durch Kapazitätsgrenzen eingeschränkt, und es ist verschwenderisch in seinen Berechnungen. Wenn es nach einer Antwort auf eine Frage sucht, erzeugt es gleichzeitig Antworten auf verwandte Fragen, und es ersetzt womöglich die erbetene Antwort durch eine Antwort, die ihm leichter einfällt. In dieser Konzeption von Heuristiken ist die heuristische Antwort nicht unbedingt einfacher oder anspruchsloser als die ursprüngliche Frage – sie ist nur leichter zugänglich, lässt sich schneller und müheloser berechnen. Die heuristischen Antworten sind nicht beliebig, und sie sind oft näherungsweise richtig. Aber manchmal liegen sie ziemlich daneben.

System 1 registriert die kognitive Leichtigkeit, mit der es Informationen verarbeitet, aber es erzeugt kein Warnsignal, wenn es unzuverlässig wird. Intuitive Antworten fallen uns schnell ein und wirken in hohem Maße überzeugend, ganz gleich, ob sie sich erworbenen Fähigkeiten oder Heuristiken verdanken. System 2 kann nicht leicht zwischen einer sachkundigen und einer heuristischen Antwort unterscheiden. Der einzige Ausweg besteht darin, einen Gang zurückzuschalten und zu versuchen, eine eigene Antwort zu konstruieren, was es nur ungern tut, weil es träge ist. Viele Vorschläge von System 1 werden ohne nähere Prüfung beiläufig unterstützt, wie etwa bei dem Schläger-und-Ball-Problem. Auf diese Weise erwirbt System 1 seinen schlechten Ruf als Quelle von Irrtümern und kognitiven Verzerrungen. Seine Funktionsmerkmale, zu denen unter anderem die WYSIATI-Regel, Intensitätsabstimmung und assoziative Kohärenz gehören, erzeugen vorhersagbare Verzerrungen und kognitive Illusionen wie etwa den Ankereffekt, nicht regressive Vorhersagen, Selbstüberschätzung und viele weitere.

Was kann man gegen Verzerrungen tun? Wie können wir Urteile und Entscheidungen verbessern, sowohl unsere eigenen als auch jene der Institutionen, denen wir dienen und die uns dienen? Kurz gesagt: Ohne erhebliche Anstrengungen kann nichts erreicht werden. Wie ich aus Erfahrung weiß, lässt sich System 1 nicht leicht erziehen. Bis auf einige Effekte, die ich überwiegend auf mein Alter zurückführe, ist mein intuitives

Denken heute noch genauso anfällig für Selbstüberschätzung, extreme Vorhersagen und den Planungsfehlschluss wie vor meinen Studien über diese Themen. Nur meine Fähigkeit, Situationen zu erkennen, in denen Fehler wahrscheinlich sind, hat sich verbessert: »Diese Zahl wird ein Anker sein …«, »Die Entscheidung könnte sich ändern, wenn das Problem neu formuliert wird …«. Und ich habe viel größere Fortschritte dabei gemacht, wenn nicht meine eigenen, dann doch die Fehler anderer zu erkennen.

Fehler, die aus System 1 hervorgehen, lassen sich prinzipiell leicht vermeiden: Man sollte die Anzeichen dafür erkennen, dass man sich in einem kognitiven Minenfeld bewegt, mental einen Gang zurückschalten und System 2 um Verstärkung bitten. So werden Sie vorgehen, wenn Sie das nächste Mal auf die Müller-Lyer-Illusion stoßen. Wenn Sie Geraden mit angesetzten Pfeilspitzen, die in verschiedene Richtungen zeigen, sehen, werden Sie die Situation als eine wiedererkennen, in der Sie Ihren Wahrnehmungen der Länge nicht trauen sollten. Leider wird diese vernünftige Vorgehensweise ausgerechnet dann am wenigsten angewandt, wenn sie am dringendsten notwendig wäre. Wir alle hätten gern eine Warnglocke, die immer dann laut läutet, wenn wir im Begriff sind, einen schwerwiegenden Fehler zu begehen, aber eine solche Glocke gibt es nicht, und kognitive Illusionen sind im Allgemeinen schwerer zu erkennen als Wahrnehmungstäuschungen. Die Stimme der Vernunft mag viel leiser sein als die laute und deutliche Stimme einer fehlerhaften Intuition, und es ist unangenehm, seine Intuitionen zu hinterfragen, wenn man dem Stress einer weitreichenden Entscheidung ausgesetzt ist. Das Letzte, was man will, wenn man in Schwierigkeiten ist, sind noch mehr Zweifel. Dies führt dazu, dass man ein Minenfeld viel leichter erkennen kann, wenn man anderen dabei zusieht, wie sie in es hineinlaufen, als wenn man es selbst tut. Beobachter sind kognitiv weniger stark beansprucht und offener für Informationen als Handelnde. Das war mein Grund dafür, ein Buch zu schreiben, das sich eher an Kritiker und Klatschmäuler wendet als an Entscheidungsträger.

Organisationen können Fehler besser vermeiden als Individuen, weil sie naturgemäß langsamer denken und die Macht haben, geordnete Abläufe durchzusetzen. Organisationen können nützliche Prüflisten erstellen und ihre Einhaltung vorschreiben, und sie können auch ausgeklügeltere Methoden wie Referenzklassen-Vorhersagen und Prä-mortem-Analysen einführen.[6] Organisationen können zumindest teilweise auch dadurch, dass

sie ein ganz bestimmtes Vokabular verwenden, eine Kultur fördern, in der Menschen aufeinander aufpassen, wenn sie sich Minenfeldern nähern.[7] Welche Produkte oder Leistungen eine Organisation auch immer bereitstellt, sie ist jedenfalls auch eine Fabrik, die Urteile und Entscheidungen produziert. Jede Fabrik ist auf Verfahren angewiesen, die die Qualität ihrer Produkte vom anfänglichen Design über die Fertigung bis zur Warenausgangskontrolle sicherstellen. Die entsprechenden Etappen in der Produktion von Entscheidungen sind das Framing des zu lösenden Problems, die Sammlung relevanter Informationen, die zu einer Entscheidung führen, Nachdenken und Überprüfen. Eine Organisation, die ihr »Entscheidungsprodukt« verbessern will, sollte routinemäßig in jeder dieser Etappen nach Effizienzverbesserungen Ausschau halten. Das operative Konzept ist Routine. Konstante Qualitätskontrolle ist eine Alternative zu umfassenden Überprüfungen der Prozesse, die Organisationen im Gefolge von Katastrophen üblicherweise durchführen. Es bleibt viel zu tun, um die Entscheidungsfindung zu verbessern. Ein Beispiel von vielen ist die bemerkenswerte Tatsache, dass die so wichtige Fähigkeit, Sitzungen effizient zu führen, nicht systematisch geschult wird.

Letztlich ist eine differenzierte Sprache von zentraler Bedeutung für die Fähigkeit zu konstruktiver Kritik. Die Identifizierung von Urteilsfehlern ist eine diagnostische Aufgabe, die ein präzises Vokabular erfordert, ganz ähnlich wie in der Medizin. Der Name einer Krankheit ist ein Haken, an dem alles, was man über diese Krankheit weiß, wie Anfälligkeiten, Umweltfaktoren, Symptome, Prognose und Therapie, aufgehängt wird. In ähnlicher Weise bringen Bezeichnungen wie »Ankereffekte«, »enges Framing« oder »überzogene Kohärenz« im Gedächtnis alles zusammen, was wir über eine kognitive Verzerrung, ihre Ursachen, ihre Wirkungen und Möglichkeiten zur Prävention wissen.

Es besteht ein direkter Zusammenhang zwischen präziserem Geplauder am Kaffeeautomaten und besseren Entscheidungen. Manchmal können sich Entscheidungsträger die Stimmen gegenwärtiger Klatschmäuler und künftiger Kritiker besser vorstellen, als sie die zögerliche Stimme ihrer eigenen Zweifel hören können. Sie werden bessere Entscheidungen treffen, wenn sie darauf vertrauen, dass ihre Kritiker intellektuell anspruchsvoll und fair sind, und wenn sie erwarten, dass ihre Entscheidung danach beurteilt wird, wie sie zustande kam, und nicht nur danach, was dabei herauskam.

# ANHANG

# *Urteile unter Unsicherheit: Heuristiken und kognitive Verzerrungen*[1]

Amos Tversky und Daniel Kahneman

Viele Entscheidungen basieren auf Überzeugungen, die die Wahrscheinlichkeit unsicherer Ereignisse betreffen, wie etwa der Ausgang einer Wahl, die Schuld eines Beklagten oder der zukünftige Wert des Dollars. Diese Überzeugungen werden für gewöhnlich ausgedrückt in Form von Aussagen wie »Ich glaube, dass …«, »Es ist möglich …«, »Es ist unwahrscheinlich, dass …« und so weiter. Gelegentlich werden Überzeugungen in Bezug auf unsichere Ereignisse in numerischer Form als Odds oder subjektive Wahrscheinlichkeiten ausgedrückt. Was bestimmt solche Überzeugungen? Wie beurteilen Menschen die Wahrscheinlichkeit eines unsicheren Ereignisses oder den Wert einer unbestimmten Größe? Dieser Artikel zeigt, dass sich Menschen auf eine begrenzte Anzahl heuristischer Prinzipien stützen, die komplexe Aufgaben der Abschätzung von Wahrscheinlichkeiten und der Vorhersage von Werten auf einfachere Urteilsoperationen reduzieren. Im Allgemeinen sind diese Heuristiken recht nützlich, aber manchmal führen sie zu schwerwiegenden, systematischen Fehlern.

Die subjektive Schätzung von Wahrscheinlichkeiten gleicht der subjektiven Schätzung physikalischer Maßangaben wie Entfernung oder Größe. All diese Urteile stützen sich auf Daten begrenzter Gültigkeit, die nach heuristischen Regeln verarbeitet werden. So wird beispielsweise die scheinbare Entfernung eines Gegenstandes teilweise nach seiner Deutlichkeit beurteilt. Je schärfer der Gegenstand gesehen wird, je näher scheint er zu sein. Diese Regel besitzt eine gewisse Gültigkeit, weil entferntere Gegenstände in jeder beliebigen Ansicht weniger deutlich wahrgenommen werden als nähere Gegenstände. Aber die Anwendung dieser Regel führt zu systematischen Fehlern bei der Entfernungsschätzung. Entfernungen werden insbesondere überschätzt, wenn die Sichtverhältnisse schlecht sind, weil dann die Konturen von Gegenständen verwischt sind. Andererseits werden Entfernungen bei guter Sicht oftmals unterschätzt,

weil Gegenstände scharf gesehen werden. Daher führt es zu weitverbreiteten Fehleinschätzungen, wenn die Deutlichkeit eines Gegenstandes als Hinweis auf seine Entfernung benutzt wird. Solche Verzerrungen finden sich auch bei intuitiven Wahrscheinlichkeitsurteilen. In diesem Aufsatz werden drei Heuristiken beschrieben, die zur Schätzung von Wahrscheinlichkeiten und zur Vorhersage von Werten benutzt werden. Die Verzerrungen, zu denen diese Heuristiken führen, werden im Einzelnen benannt, und die praktischen und theoretischen Folgen dieser Beobachtungen werden diskutiert.

## Repräsentativität

Viele der probabilistischen Fragen, mit denen sich Menschen befassen, gehören zu einem der folgenden Typen: Wie hoch ist die Wahrscheinlichkeit, dass Objekt A zur Klasse B gehört? Wie hoch ist die Wahrscheinlichkeit, dass Ereignis A seinen Ursprung in Prozess B hat? Wie hoch ist die Wahrscheinlichkeit, dass Prozess B Ereignis A hervorbringt? Bei der Beantwortung derartiger Fragen stützen sich Menschen typischerweise auf die Repräsentativitätsheuristik, bei der Wahrscheinlichkeiten nach dem Ausmaß, in dem A repräsentativ für B ist, beurteilt werden, also nach dem Ausmaß, in dem A B ähnelt. Wenn beispielsweise A in hohem Maße repräsentativ für B ist, dann wird die Wahrscheinlichkeit als hoch eingeschätzt, dass A aus B hervorgegangen ist. Besitzt andererseits A keine Ähnlichkeit mit B, wird die Wahrscheinlichkeit, dass A aus B hervorgegangen ist, als niedrig eingestuft.

Betrachten wir zur Verdeutlichung eines Urteils auf der Basis der Repräsentativität ein Individuum, das von einem ehemaligen Nachbarn folgendermaßen beschrieben wurde: »Steve ist sehr schüchtern und reserviert, stets hilfsbereit, aber wenig an Menschen oder der realen Welt interessiert. Als ein sanftmütiger und penibler Mensch hat er ein Bedürfnis nach Ordnung und Struktur und eine Passion für Details.« Wie hoch schätzen Menschen die Wahrscheinlichkeit ein, dass Steve in einem bestimmten Beruf aus einer Liste von Wahlmöglichkeiten (zum Beispiel Landwirt, Verkäufer, Zivilpilot, Bibliothekar oder Arzt) tätig ist? Wie ordnen die Befragten diese Berufstätigkeiten von den wahrscheinlichsten zu den unwahrscheinlichsten Möglichkeiten? Bei der Repräsentativitätsheuristik wird die Wahrscheinlichkeit, dass Steve zum Beispiel Bibliothekar ist, nach dem Ausmaß beurteilt, in dem er repräsentativ für das Stereotyp eines Bibliothekars ist, diesem also ähnelt. Tatsächlich haben Studien über derartige Probleme gezeigt, dass sich die Rangordnung der Berufstätigkeiten nach Wahrscheinlichkeit exakt mit der Rangordnung nach Ähnlichkeit deckt.[2] Diese Methode, Wahrscheinlichkeiten zu beurteilen, führt zu schwerwiegenden Fehlern, weil Ähnlichkeit oder Repräsentativität von einigen Faktoren, die sich auf Wahrscheinlichkeitsurteile auswirken sollten, nicht beeinflusst wird.

*Unempfindlichkeit gegen A-priori-Wahrscheinlichkeiten der Ergebnisse.* Einer der Faktoren, die keinen Einfluss auf die Repräsentativität haben, sich aber deutlich auf die Wahrscheinlichkeit auswirken sollten, ist die A-priori-Wahrscheinlichkeit oder Grundhäufigkeit (Basisrate) der Ergebnisse. Im Fall von Steve beispielsweise sollte die Tatsache, dass es in der Bevölkerung viel mehr Landwirte als Bibliothekare gibt, in jede vernünftige Schätzung der Wahrscheinlichkeit eingehen, dass Steve eher für einen Bibliothekar als für einen Landwirt gehalten wird. Die Berücksichtigung der Grundhäufigkeit wirkt sich allerdings nicht auf die Ähnlichkeit von Steve mit den Stereotypen von Bibliothekaren und Landwirten aus. Wenn Menschen Wahrscheinlichkeiten anhand der Repräsentativität beurteilen, werden sie daher die A-priori-Wahrscheinlichkeiten vernachlässigen. Diese Hypothese wurde in einem Experiment überprüft, in dem die A-priori-Wahrscheinlichkeiten manipuliert wurden.[3] Den Versuchsteilnehmern wurden kurze Persönlichkeitsbeschreibungen mehrerer Individuen dargeboten, die angeblich aus einer Gruppe von hundert Akademikern – Ingenieuren und Juristen – nach dem Zufallsprinzip ausgewählt worden waren. Die Probanden sollten für jede Beschreibung die Wahrscheinlichkeit angeben, mit der sie sich auf einen Ingenieur oder einen Juristen bezog. In einer Variante des Experiments wurde den Versuchspersonen gesagt, die Gruppe, aus der die Beschreibungen entnommen worden seien, habe aus siebzig Ingenieuren und dreißig Juristen bestanden. In einer anderen Variante wurde den Probanden gesagt, die Gruppe habe aus dreißig Ingenieuren und siebzig Juristen bestanden. Die Wahrscheinlichkeit (Odds), dass sich eine bestimmte Beschreibung auf einen Ingenieur statt auf einen Juristen bezieht, sollte in der ersten Bedingung höher sein – denn dort stellen Ingenieure die Mehrheit – als in der zweiten Bedingung, in der Juristen die Mehrheit stellen. Durch Anwendung der Bayesschen Regel kann insbesondere gezeigt werden, dass das Verhältnis dieser Odds für jede Beschreibung $(0{,}70/0{,}30)^2$ oder 5,44 betragen sollte. In einem eklatanten Verstoß gegen die Bayessche Regel trafen die Versuchspersonen in den beiden Bedingungen im Wesentlichen die gleichen Wahrscheinlichkeitsurteile. Offenbar beurteilten die Testpersonen die Wahrscheinlichkeit, dass sich eine bestimmte Beschreibung auf einen Ingenieur statt auf einen Juristen bezieht, nach dem Ausmaß, in dem diese

Beschreibung repräsentativ für die beiden Stereotype ist, wobei sie den A-priori-Wahrscheinlichkeiten der Kategorien nur eine geringe oder gar keine Beachtung schenkten.

Die Probanden verwendeten die A-priori-Wahrscheinlichkeiten richtig, wenn sie sonst keine Informationen hatten. Wenn ihnen keine Persönlichkeitsskizze vorlag, schätzten sie die Wahrscheinlichkeit, dass ein unbekanntes Individuum ein Ingenieur ist, unter beiden experimentellen Bedingungen mit je zwei Basisraten auf 0,70 beziehungsweise 0,30. Doch wurden die A-priori-Wahrscheinlichkeiten faktisch ignoriert, sobald eine Beschreibung vorgelegt wurde, auch wenn diese Beschreibung keinerlei Informationsgehalt hatte. Die Antworten auf die folgende Beschreibung veranschaulichen dieses Phänomen:

> Dick ist ein dreißigjähriger Mann. Er ist verheiratet, aber kinderlos. Als ein äußerst fähiger, hoch motivierter Mann verspricht er in seinem Berufsfeld einmal recht erfolgreich zu sein. Bei seinen Kollegen ist er sehr beliebt.

Diese Beschreibung sollte keine Informationen hinsichtlich der Frage liefern, ob Dick Ingenieur oder Jurist ist. Folglich sollte die Wahrscheinlichkeit, dass Dick ein Ingenieur ist, gleich dem Prozentsatz der Ingenieure in der Gruppe sein, so, als wäre keine Beschreibung gegeben worden. Doch die Probanden schätzten die Wahrscheinlichkeit, dass Dick ein Ingenieur ist, auf 0,50, unabhängig davon, ob der angegebene Prozentsatz der Ingenieure in der Gruppe 70 oder 30 Prozent war. Offenkundig reagierten Menschen unterschiedlich, je nachdem, ob sie keine oder wertlose Informationen erhielten. Ohne Angabe spezifischer Informationen werden die A-priori-Wahrscheinlichkeiten richtig angewandt; werden wertlose Informationen angegeben, werden die A-priori-Wahrscheinlichkeiten ignoriert.[4]

*Unempfindlichkeit gegenüber dem Stichprobenumfang.* Um die Wahrscheinlichkeit für ein bestimmtes Ergebnis in einer Stichprobe abzuschätzen, die einer näher beschriebenen Grundgesamtheit entnommen

wurde, greifen Menschen typischerweise auf die Repräsentativitätsheuristik zurück. Das heißt, sie beurteilen die Wahrscheinlichkeit eines Stichprobenergebnisses – zum Beispiel, dass die durchschnittliche Körpergröße in einer Zufallsstichprobe von zehn Männern 1,83 Meter beträgt – anhand der Ähnlichkeit dieses Ergebnisses mit dem entsprechenden Parameter (das heißt der durchschnittlichen Körpergröße in der Grundgesamtheit der Männer). Die Ähnlichkeit einer Stichprobenstatistik mit einem Populationsparameter hängt nicht von dem Umfang der Stichprobe ab. Wenn Wahrscheinlichkeiten nach der Repräsentativität beurteilt werden, ist folglich die beurteilte Wahrscheinlichkeit einer Stichprobenstatistik weitgehend unabhängig vom Stichprobenumfang. Als Probanden die Verteilungen der Durchschnittsgröße für Stichproben von unterschiedlichen Umfängen beurteilten, erzeugten sie identische Verteilungen. So wurde zum Beispiel der Wahrscheinlichkeit, eine Durchschnittsgröße von über 1,83 Meter zu erreichen, für Stichproben aus tausend, hundert und zehn Männern derselbe Wert zugeschrieben. [5] Außerdem haben Probanden die Rolle des Stichprobenumfangs auch dann nicht angemessen berücksichtigt, wenn diese in der Formulierung des Problems betont wurde. Betrachten Sie die folgende Frage:

> In einer bestimmten Stadt gibt es zwei Kliniken. In der größeren Klinik kommen jeden Tag etwa 45 Kinder zur Welt, in der kleineren Klinik werden jeden Tag etwa 15 Kinder geboren. Wie Sie wissen, sind etwa 50 Prozent aller Neugeborenen Jungen. Die genauen Prozentsätze schwanken allerdings von Tag zu Tag. Manchmal mögen es mehr als 50 Prozent sein, manchmal weniger. Über einen Zeitraum von einem Jahr vermerkte jede Klinik die Tage, an denen über 60 Prozent der Neugeborenen Jungen waren. Welche Klinik verzeichnete Ihrer Meinung nach mehr von diesen Tagen?

Die größere Klinik (21)

Die kleinere Klinik (21)

Beide etwa gleich

(das heißt mit einer Differenz von höchsten 5 Prozent) (53)

Die Werte in Klammern sind die Zahlen der Studenten, die die jeweilige Antwort wählten.

Die meisten Probanden beurteilten die Wahrscheinlichkeit, dass über 60 Prozent der Neugeborenen Jungen sind, in der kleinen und der großen Klinik als gleich groß, vermutlich weil diese Ereignisse durch die gleiche Statistik beschrieben werden und daher gleichermaßen repräsentativ für die allgemeine Bevölkerung sind. Dagegen folgt aus der Stichprobentheorie, dass die erwartete Anzahl von Tagen, an denen über 60 Prozent der Neugeborenen Jungen sind, in der kleinen Klinik viel größer ist als in der großen, weil eine große Stichprobe mit geringerer Wahrscheinlichkeit von 50 Prozent abweicht. Diese fundamentale statistische Tatsache gehört offenkundig nicht zum Repertoire menschlicher Intuitionen.

Eine ähnliche Unempfindlichkeit gegenüber dem Stichprobenumfang wurde bei Urteilen über die A-posteriori-Wahrscheinlichkeit berichtet, also der Wahrscheinlichkeit, dass eine Stichprobe einer bestimmten Grundgesamtheit und nicht einer anderen entnommen wurde. Betrachten wir das folgende Beispiel:

> Stellen Sie sich eine mit Kugeln gefüllte Urne vor; zwei Drittel der Kugeln haben eine Farbe, und ein Drittel hat eine andere Farbe. Eine Person hat fünf Kugeln aus der Urne gezogen – vier rote und eine weiße. Eine andere Person hat zwanzig Kugeln gezogen – zwölf rote und acht weiße. Welche der beiden Personen sollte sicherer sein, dass die Urne zwei Drittel rote Kugeln und ein Drittel weiße Kugeln enthält statt des

Gegenteils? Welche Wahrscheinlichkeit sollte jede Person angeben?

Bei diesem Problem sind die richtigen A-posteriori-Wahrscheinlichkeiten 8 zu 1 für die 4:1-Stichprobe und 16 zu 1 für die 12:8-Stichprobe, unter der Annahme gleicher A-priori-Wahrscheinlichkeiten. Die meisten Menschen glauben jedoch, die erste Stichprobe liefere viel stärkere Belege für die Hypothese, dass die Urne überwiegend rote Kugeln enthält, weil der Anteil der roten Kugeln in der ersten Stichprobe größer ist als in der zweiten. Auch hier werden intuitive Urteile von den Stichprobenanteilen bestimmt, während sie weitgehend unbeeinflusst sind vom Stichprobenumfang, der eine entscheidende Rolle bei der Bestimmung der tatsächlichen A-posteriori-Wahrscheinlichkeiten spielt.[6] Außerdem sind intuitive Schätzungen der A-posteriori-Wahrscheinlichkeiten weit weniger extrem als die richtigen Werte. Bei Problemen dieses Typs wurde wiederholt beobachtet, dass der Einfluss empirischer Daten unterschätzt wurde.[7] Dieses Phänomen wurde »Konservatismus« genannt.

*Falsche Zufallskonzepte.* Menschen erwarten, dass eine Folge von Ereignissen, die durch einen Zufallsprozess generiert wird, die wesentlichen Merkmale dieses Prozesses auch dann repräsentiert, wenn die Folge kurz ist. Wenn Menschen zum Beispiel Münzwürfe danach beurteilen, wie oft »Kopf« oder »Zahl« auftreten, stufen sie die Folge K-Z-K-Z-Z-K als wahrscheinlicher ein als die Folge K-K-K-Z-Z-Z, die nicht zufällig zu sein scheint, und auch als wahrscheinlicher als die Folge K-K-K-K-Z-K, die nicht die Fairness der Münze repräsentiert. [8] Menschen erwarten also, dass die wesentlichen Merkmale repräsentiert werden, nicht nur global in der gesamten Folge, sondern auch lokal in jedem ihrer Abschnitte. Aber eine lokal repräsentative Folge weicht systematisch von der Zufallserwartung ab: Sie enthält zu viele Wechsel und zu wenig Reihen. Eine weitere Folge des Glaubens an die lokale Repräsentativität ist der bekannte Spielerfehlschluss. Nachdem sie eine lange Reihe von Rot beim Roulette beobachtet haben, glauben die meisten Menschen irrigerweise, jetzt sei Schwarz fällig, vermutlich weil das Auftreten von Schwarz eine repräsentativere Folge hervorbringen wird als ein weiteres Auftreten von

Rot. Der Zufall wird im Allgemeinen als ein Prozess der Selbstkorrektur angesehen, bei dem eine Abweichung in eine Richtung eine Abweichung in die andere Richtung auslöst, um das Gleichgewicht wiederherzustellen. Tatsächlich werden Abweichungen nicht in dem Maße »korrigiert«, wie sich der Zufallsprozess entfaltet, sie werden lediglich verwässert.

Falsche Zufallskonzeptionen sind nicht auf naive Probanden beschränkt. Eine Studie über die statistischen Intuitionen von erfahrenen forschenden Psychologen [9] enthüllte einen nachhaltigen Glauben in das, was man das »Gesetz der kleinen Zahlen« nennen könnte, dem zufolge auch kleine Stichproben höchst repräsentativ für die Grundgesamtheiten sind, denen sie entnommen wurden. In den Antworten dieser Forscher spiegelte sich die Erwartung wider, dass eine gültige Hypothese über eine Population durch ein statistisch signifikantes Ergebnis in einer Stichprobe repräsentiert wird, deren Größe nicht weiter von Belang wäre. Infolgedessen überbewerten die Forscher die Aussagekraft von Ergebnissen kleiner Stichproben, und sie überschätzen in erheblichem Umfang die Replizierbarkeit solcher Ergebnisse. Bei der praktischen Durchführung von Forschungsprojekten bewirkt diese Verzerrung, dass Stichproben unzulänglicher Größe ausgewählt und Studienergebnisse überinterpretiert werden.

*Unempfindlichkeit für Vorhersagbarkeit.* Menschen sollen manchmal numerische Vorhersagen machen, etwa über den zukünftigen Kurs einer Aktie, die Nachfrage nach einer Ware oder den Ausgang eines Footballspiels. Solche Vorhersagen erfolgen oftmals auf der Basis der Repräsentativität. Nehmen wir zum Beispiel an, jemand erhält eine Beschreibung eines Unternehmens und er soll nun dessen zukünftigen Gewinn vorhersagen. Wenn das Unternehmen sehr positiv beschrieben wird, wird ein sehr hoher Gewinn am repräsentativsten für diese Beschreibung erscheinen. Weder die Wirklichkeitstreue der Beschreibung noch das Ausmaß, in dem sie zutreffende Vorhersagen erlaubt, wirken sich darauf aus, wie positiv die Beschreibung ist. Wenn Menschen daher Vorhersagen ausschließlich auf der Basis der Vorteilhaftigkeit der Beschreibung machen, werden sich die Zuverlässigkeit der empirischen Daten und die erwartete Vorhersagegenauigkeit nicht auf Vorhersagen auswirken.

Diese Form der Urteilsbildung verstößt gegen die normative statistische Theorie, in der die Extremität und die Spannweite der Vorhersagen durch

Vorhersagbarkeitserwägungen kontrolliert werden. Wenn die Vorhersagbarkeit gleich null ist, sollte in allen Fällen die gleiche Vorhersage gemacht werden. Wenn die Beschreibungen von Unternehmen zum Beispiel keine Informationen enthalten, die für den Gewinn von Bedeutung sind, dann sollte für alle Unternehmen derselbe Wert (wie etwa der durchschnittliche Gewinn) vorhergesagt werden. Bei vollkommener Vorhersagbarkeit werden die vorhergesagten Werte mit den tatsächlichen Werten übereinstimmen, und die Spannweite der Vorhersagen wird sich mit der Spannweite der Ergebnisse decken. Im Allgemeinen gilt: Je höher die Vorhersagbarkeit, desto größer die Spannweite der vorhergesagten Werte.

Mehrere Studien über numerische Vorhersage haben gezeigt, dass intuitive Vorhersagen diese Regel verletzen und dass Probanden Vorhersagbarkeitserwägungen keine oder nur eine geringe Beachtung schenken.[10] In einer dieser Studien wurden Probanden mehrere Absätze Text dargeboten, die jeweils die Leistung eines Lehramtsreferendars in einer Lehrprobe beschrieben. Einige Probanden wurden aufgefordert, die Qualität der in dem Absatz beschriebenen Unterrichtsstunde im Verhältnis zu einer näher beschriebenen Population in Prozenträngen (Perzentil-Scores) zu bewerten. Andere Probanden sollten, ebenfalls in Prozenträngen, die Stellung jedes Referendars fünf Jahre nach der Lehrprobe vorhersagen. In beiden Bedingungen waren die Urteile gleich. Das heißt, die Vorhersage eines zukünftigen Kriteriums (Erfolg eines Lehrers nach fünf Jahren) war identisch mit der Bewertung der Informationen, auf denen die Vorhersage beruhte (die Qualität der Lehrprobe). Die Studenten, die diese Vorhersagen machten, waren sich zweifellos der beschränkten Vorhersagbarkeit der pädagogischen Kompetenz auf der Basis einer einzigen Lehrprobe fünf Jahre zuvor bewusst; trotzdem hatten ihre Vorhersagen die gleiche Spannweite wie ihre Bewertungen.

*Die Illusion der Gültigkeit.* Wie wir sahen, verfahren Menschen bei Vorhersagen oftmals so, dass sie jenes Ergebnis (zum Beispiel eine Berufstätigkeit) auswählen, das am repräsentativsten für den Input (zum Beispiel die Beschreibung einer Person) ist. Das Ausmaß, in dem sie von der Richtigkeit ihrer Vorhersage überzeugt sind, hängt in erster Linie von dem Grad der Repräsentativität ab (das heißt von der Qualität der Übereinstimmung zwischen dem ausgewählten Ergebnis und dem Input), wobei den Faktoren, welche die Vorhersagegenauigkeit einschränken, nur

eine geringe oder gar keine Beachtung geschenkt wird. Das heißt, Menschen geben an, fest von der Vorhersage überzeugt zu sein, dass eine Person ein Bibliothekar ist, wenn sie eine Beschreibung seiner Persönlichkeit erhalten, die dem Stereotyp von Bibliothekaren entspricht, selbst wenn die Beschreibung knapp, unzuverlässig oder veraltet ist. Die ungerechtfertigte subjektive Überzeugung (von der Richtigkeit der Vorhersage), die durch eine gute Übereinstimmung zwischen dem vorhergesagten Ergebnis und der Input-Information erzeugt wird, könnte »Illusion der Gültigkeit« genannt werden. Diese Illusion besteht auch dann fort, wenn der Urteilende sich der Faktoren bewusst ist, die die Genauigkeit seiner Vorhersagen einschränken. Es ist eine weitverbreitete Beobachtung, dass Psychologen, die Auswahlinterviews führen, oftmals in hohem Maße von der Richtigkeit ihrer Vorhersagen überzeugt sind, selbst wenn sie die riesige Anzahl von Publikationen kennen, die zeigen, dass Auswahlinterviews in hohem Maße fehleranfällig sind. Die fortdauernde Anwendung des klinischen Interviews zu Auswahlzwecken – trotz wiederholter Nachweise seiner Mängel – bestätigt lediglich die Stärke dieses Effekts.

Die innere Konsistenz eines Eingabemusters ist eine wichtige Bestimmungsgröße der subjektiven Überzeugung von der Richtigkeit der Vorhersagen auf der Basis dieser Inputs. So geben Menschen an, sie seien sich bei ihrer Vorhersage des abschließenden Notendurchschnitts eines Studenten, der im ersten Studienjahr nur Zweien bekam, ihrer Sache sicherer, als wenn sie den Notendurchschnitt eines Studenten vorhersagten, der im ersten Jahr zahlreiche Einser und Dreier bekam. Hochkonsistente Muster werden am häufigsten dann beobachtet, wenn die Input-Variablen in hohem Maße redundant oder korreliert sind. Daher sind Menschen tendenziell fest von Vorhersagen überzeugt, die auf redundanten Input-Variablen basieren. Ein elementares Resultat in der Statistik der Korrelation besagt jedoch, dass bei gegebenen Input-Variablen ausgewiesener Gültigkeit eine Vorhersage auf der Basis mehrerer solcher Inputs eine höhere Treffgenauigkeit erreichen kann, wenn diese unabhängig voneinander sind, als wenn sie redundant oder korreliert sind. Die Redundanz der Eingabedaten vermindert daher die Vorhersagegenauigkeit, während sie zugleich die subjektive Überzeugung von der Richtigkeit der Vorhersagen erhöht, und Menschen sind oftmals fest von Vorhersagen überzeugt, die mit hoher Wahrscheinlichkeit danebenliegen.[11]

*Falsches Verständnis der statistischen Regression.* Angenommen, eine große Gruppe von Kindern wurde mit zwei äquivalenten Versionen eines Befähigungstests geprüft. Wenn man unter denjenigen, die in einer der beiden Versionen am besten abschnitten, zehn Kinder auswählt, wird ihre Leistung in der zweiten Version für gewöhnlich ein wenig enttäuschend sein. Wenn man, umgekehrt, unter denjenigen, die in der ersten Testversion am schlechtesten abschnitten, zehn Kinder auswählt, werden diese in der anderen Version im Schnitt etwas besser abschneiden. Betrachten wir nun, allgemeiner, die beiden Variablen X und Y, die die gleiche Verteilung haben. Wenn man Individuen auswählt, deren durchschnittlicher X-Score um k Einheiten von dem Mittelwert von X abweicht, dann wird der Durchschnitt ihrer Y-Scores um weniger als k Einheiten von dem Mittelwert von Y abweichen. Diese Beobachtungen veranschaulichen ein allgemeines Phänomen, das als »Regression zum Mittelwert« bezeichnet wird und das vor über hundert Jahren erstmals von Galton beschrieben wurde.

Im normalen Lauf des Lebens begegnet man vielen Beispielen einer Regression zum Mittelwert, etwa beim Vergleich der Körpergröße von Vätern und Söhnen, der Intelligenz von Ehemännern und -frauen oder der Leistung von Individuen bei aufeinanderfolgenden Prüfungen. Trotzdem entwickeln Menschen keine zutreffenden Intuitionen über dieses Phänomen. Erstens rechnen sie in vielen Kontexten, in denen eine solche Regression praktisch vorprogrammiert ist, nicht damit. Zweitens, wenn sie das Auftreten einer Regression bemerken, denken sie sich oftmals fadenscheinige kausale Erklärungen dafür aus.[12] Wir behaupten, dass das Phänomen der Regression schwer fassbar bleibt, weil es unvereinbar ist mit der Überzeugung, dass das vorhergesagte Ergebnis maximal repräsentativ für den Input sein sollte und dass daher der Wert der Ergebnis-Variable genauso extrem sein sollte wie der Wert der Eingabe-Variable.

Wird die Tragweite des Phänomens der Regression verkannt, kann dies weitreichende negative Folgen haben, wie die folgende Beobachtung verdeutlicht. [13] In einer Diskussion über Flugtraining wiesen erfahrene Fluglehrer darauf hin, dass auf das Lob für eine außergewöhnlich sanfte Landung typischerweise eine schlechtere Landung beim nächsten Versuch folgt, während auf scharfe Kritik nach einer holprigen Landung für gewöhnlich eine Verbesserung beim nächsten Versuch folgt. Die Fluglehrer gelangten zu dem Schluss, dass – im Gegensatz zur herrschenden

psychologischen Lehrmeinung – verbales Lob dem Lernen abträglich sei, während verbaler Tadel dem Lernen förderlich sei. Weil auch hier die Regression zum Mittelwert am Werk ist, ist diese Schlussfolgerung nicht gerechtfertigt. Wie in anderen Fällen wiederholter Überprüfung wird auf eine schlechte Leistung in der Regel eine Verbesserung folgen, während auf eine hervorragende Leistung für gewöhnlich eine Verschlechterung folgen wird, auch wenn der Ausbilder nicht auf die Leistung des Auszubildenden beim ersten Versuch reagiert. Weil die Ausbilder ihre Schüler nach guten Landungen lobten und nach schlechten Landungen tadelten, gelangten sie zu der irrigen und potenziell schädlichen Schlussfolgerung, Tadel sei wirkungsvoller als Lob.

Mangelndes Verständnis für den Regressionseffekt veranlasst einen dazu, die Effektivität von Bestrafungen zu überschätzen und die Effektivität von Belohnungen zu unterschätzen. In sozialen Interaktionen und bei Schulungen werden gute Leistungen typischerweise belohnt, während schlechte Leistungen in der Regel bestraft werden. Allein regressionsbedingt wird sich das Verhalten daher mit hoher Wahrscheinlichkeit nach einer Bestrafung verbessern und nach einer Belohnung verschlechtern. Folglich ist die *Conditio humana* so beschaffen, dass man allein aufgrund des Zufalls meistens dafür belohnt wird, dass man andere bestraft, und meistens dafür bestraft, dass man sie belohnt. Die Menschen sind sich dieser Kontingenz im Allgemeinen nicht bewusst. Tatsächlich scheint der schwer fassbare Einfluss der Regression auf die scheinbaren Folgen von Belohnung und Bestrafung der Aufmerksamkeit der Wissenschaftler auf diesem Gebiet entgangen zu sein.

## Verfügbarkeit

Es gibt Situationen, in denen Menschen die Häufigkeit einer Klasse oder die Wahrscheinlichkeit eines Ereignisses nach der Leichtigkeit beurteilen, mit der ihnen Beispiele oder Fälle einfallen. So mag man zum Beispiel das Herzinfarktrisiko bei Menschen mittleren Alters danach beurteilen, an wie viele entsprechende Fälle in seinem Bekanntenkreis man sich erinnert. In ähnlicher Weise mag man die Wahrscheinlichkeit, dass ein bestimmtes Geschäftsprojekt scheitert, dadurch beurteilen, dass man sich die verschiedenen Schwierigkeiten ausmalt, auf die es stoßen könnte. Diese Urteilsheuristik wird »Verfügbarkeit« genannt. Die Verfügbarkeit liefert einen nützlichen Anhaltspunkt für die Beurteilung der Häufigkeit oder Wahrscheinlichkeit, weil Beispielfälle aus großen Klassen in der Regel besser und schneller abgerufen werden als Beispiele aus weniger häufigen Klassen. Doch wird die Verfügbarkeit von anderen Faktoren als der Häufigkeit und Wahrscheinlichkeit beeinflusst. Folglich führt die Orientierung an der Verfügbarkeit zu vorhersagbaren Verzerrungen, von denen einige nachfolgend veranschaulicht werden.

*Verzerrungen, die auf die Abrufbarkeit von Beispielen zurückzuführen sind.* Wenn die Größe einer Klasse nach der Verfügbarkeit ihrer Beispielfälle beurteilt wird, erscheint eine Klasse, deren Beispielfälle sich leicht abrufen lassen, größer als eine Klasse gleicher Häufigkeit, deren Beispiele nicht so leicht abrufbar sind. Bei einer einfachen Demonstration dieses Effekts hörten Testpersonen eine Liste bekannter Persönlichkeiten beider Geschlechter, und sie sollten anschließend einschätzen, ob die Liste mehr Namen von Männern als von Frauen enthielt. Verschiedenen Probandengruppen wurden verschiedene Listen dargeboten. Auf einigen der Listen waren die Männer vergleichsweise berühmter als die Frauen, während auf anderen die Frauen vergleichsweise berühmter waren als die Männer. Bei jeder der Listen urteilten die Versuchspersonen fälschlicherweise, dass die Klasse (Geschlecht) mit mehr berühmten Persönlichkeiten die größere sei.[14]

Neben der Vertrautheit gibt es weitere Faktoren, etwa die Salienz (Auffälligkeit), die die Abrufbarkeit von Beispielen beeinflussen. So hat

zum Beispiel der Anblick eines brennenden Hauses vermutlich einen größeren Einfluss auf die subjektive Wahrscheinlichkeit solcher Vorfälle als das Lesen eines Artikels über einen Brand in einer lokalen Zeitung. Außerdem sind Beispiele aus jüngerer Vergangenheit vergleichsweise verfügbarer als ältere Beispiele. Es ist eine weitverbreitete Erfahrung, dass die subjektive Wahrscheinlichkeit von Verkehrsunfällen vorübergehend ansteigt, wenn man ein umgekipptes Auto auf dem Seitenstreifen sieht.

*Verzerrungen, die auf die Effektivität einer Suchmenge zurückzuführen sind.* Angenommen, man wählt aus einem englischen Text nach dem Zufallsprinzip ein Wort (bestehend aus drei oder mehr Buchstaben) aus. Ist es wahrscheinlicher, dass das Wort mit r beginnt oder dass r der dritte Buchstabe ist? Menschen gehen an dieses Problem heran, indem sie sich an Wörter erinnern, die mit r beginnen *(road)*, und an Wörter, die ein r an dritter Stelle haben *(car)*, und anhand der Leichtigkeit, mit der ihnen Wörter der beiden Typen einfallen, beurteilen sie deren relative Häufigkeit. Weil es viel leichter ist, anhand des ersten als anhand des dritten Buchstabens nach Wörtern zu suchen, urteilen die meisten Menschen, dass Wörter, die mit einem bestimmten Konsonanten beginnen, häufiger sind als Wörter, in denen der gleiche Konsonant an dritter Stelle auftaucht. Sie tun dies auch bei Konsonanten wie etwa r oder k, die an dritter Stelle häufiger sind als an erster Stelle.[15]

*Verschiedene Aufgaben aktivieren unterschiedliche Suchmengen.* Nehmen wir zum Beispiel an, Sie sollen die Häufigkeit einschätzen, mit der abstrakte Wörter *(thought, love)* und konkrete Wörter *(door, water)* im geschriebenen Englisch vorkommen. Ein naheliegender Weg zur Beantwortung dieser Frage ist es, nach Kontexten zu suchen, in denen das Wort auftauchen könnte. Es scheint leichter zu sein, an Kontexte zu denken, in denen ein abstrakter Begriff erwähnt wird (»Liebe« in Liebesgeschichten), als an Kontexte, in denen ein konkretes Wort (wie etwa »Tür«) erwähnt wird. Wenn die Häufigkeit von Wörtern nach der Verfügbarkeit der Kontexte beurteilt wird, in denen sie vorkommen, dann werden abstrakte Wörter als vergleichsweise häufiger als konkrete Wörter beurteilt werden. Diese Verzerrung wurde in einer neueren Studie[16] beobachtet, die nachwies, dass die beurteilte Häufigkeit des Vorkommens

abstrakter Wörter viel höher war als die konkreter Wörter, verglichen mit der objektiven Häufigkeit. Abstrakte Wörter sollen nach dieser Einschätzung auch in einer viel größeren Vielfalt von Kontexten vorkommen als konkrete Wörter.

*Verzerrungen der Vorstellbarkeit.* Manchmal muss man die Häufigkeit einer Klasse beurteilen, deren Beispiele nicht im Gedächtnis gespeichert sind, aber gemäß einer bestimmten Regel generiert werden können. In solchen Situationen erzeugt man typischerweise mehrere Beispiele und beurteilt die Häufigkeit oder Wahrscheinlichkeit nach der Leichtigkeit, mit der die relevanten Beispiele konstruiert werden können. Doch die Leichtigkeit, mit der sich Beispiele konstruieren lassen, spiegelt nicht immer ihre tatsächliche Häufigkeit wider, und diese Art der Beurteilung ist anfällig für Verzerrungen. Um dies zu verdeutlichen, wollen wir eine Gruppe von zehn Personen betrachten, die Ausschüsse aus k Mitgliedern bilden, wobei gilt $2 \leq k \leq 8$. Wie viele verschiedene Ausschüsse aus k Mitgliedern lassen sich bilden? Die richtige Lösung für dieses Problem ist gegeben durch den Binomialkoeffizienten (10/k), der für $k = 5$ ein Maximum von 252 erreicht. Die Anzahl der Ausschüsse mit k Mitgliedern ist eindeutig gleich der Anzahl der Ausschüsse aus (10-k) Mitgliedern, weil jeder Ausschuss aus k Mitgliedern eine eindeutige Gruppe von (10-k) Nichtmitgliedern definiert.

Eine Möglichkeit, diese Frage ohne Berechnung zu beantworten, besteht darin, im Geiste Ausschüsse aus k Mitgliedern zu konstruieren und ihre Zahl nach der Leichtigkeit, mit der sie einem einfallen, zu beurteilen. Ausschüsse mit wenigen Mitgliedern, etwa zwei, sind verfügbarer als Ausschüsse mit vielen Mitgliedern, zum Beispiel acht. Der einfache Plan zur Konstruktion von Ausschüssen ist eine Aufteilung der Gruppen in disjunkte Mengen. Es ist ohne Weiteres zu ersehen, dass man leicht fünf disjunkte Ausschüsse mit je zwei Mitgliedern konstruieren kann, während es unmöglich ist, auch nur zwei disjunkte Ausschüsse mit acht Mitgliedern aufzustellen. Wenn die Häufigkeit daher nach der leichten Vorstellbarkeit beziehungsweise nach der Verfügbarkeit für die Konstruktion beurteilt wird, wird es scheinbar mehr kleine als größere Ausschüsse geben, im Gegensatz zu der richtigen glockenförmigen Funktion. Als unbefangene Probanden die Anzahl eigenständiger Ausschüsse verschiedener Größe abschätzen sollten, entsprachen ihre Schätzungen tatsächlich einer abnehmenden monotonen Funktion der Ausschussgröße.[17] So betrug der mittlere Schätzwert für die

Anzahl der Ausschüsse mit zwei Mitgliedern siebzig, während die Schätzung für Ausschüsse mit acht Mitgliedern zwanzig war (die richtige Antwort ist in beiden Fällen 45).

Die Vorstellbarkeit spielt eine wichtige Rolle bei der Beurteilung von Wahrscheinlichkeiten in realen Situationen. So wird beispielsweise das Risiko, das mit einer abenteuerlichen Expedition verbunden ist, dadurch beurteilt, dass man sich Eventualitäten vorstellt, für deren Bewältigung die Expedition nicht gerüstet ist. Wenn viele solcher Schwierigkeiten plastisch ausgemalt werden, kann der Eindruck entstehen, die Expedition wäre äußerst gefährlich, obgleich die Leichtigkeit, mit der sich Katastrophen vorstellen lassen, nicht deren tatsächliche Wahrscheinlichkeit widerspiegeln muss. Umgekehrt mag das Risiko, das mit einer Unternehmung verbunden ist, erheblich unterschätzt werden, wenn sich einige potenzielle Gefahren entweder nur schwer anschaulich vorstellen lassen oder wenn sie einem schlichtweg nicht einfallen.

*Illusorische Korrelation.* Chapman und Chapman[18] haben eine interessante Verzerrung bei der Beurteilung der Häufigkeit, mit der zwei Ereignisse gemeinsam eintreten, beschrieben. Sie präsentierten unbefangen Urteilenden Informationen, die mehrere hypothetische psychiatrische Patienten betrafen. Die Daten für jeden Patienten bestanden aus einer klinischen Diagnose und aus der Zeichnung einer Person, die der Patient angefertigt hatte. Später schätzten die Urteilenden die Häufigkeit, mit der jede Diagnose (wie etwa Paranoia oder krankhaftes Misstrauen) mit unterschiedlichen Merkmalen der Zeichnung (etwa eigentümlichen Augen) verbunden war. Die Probanden überschätzten deutlich die Häufigkeit des gemeinsamen Auftretens »natürlicher Assoziationspaare«, wie etwa Argwohn und eigentümliche Augen. Dieser Effekt wurde »illusorische Korrelation« genannt. In ihren fehlerhaften Beurteilungen der Daten, die ihnen dargeboten wurden, entdeckten naive Probanden einen Großteil der weitverbreiteten, aber unbegründeten klinischen Überzeugungen über die Interpretation des *Draw-A-Person Test* (DAP). Der Effekt der illusorischen Korrelation war extrem resistent gegen widersprechende Daten. Er blieb auch dann bestehen, wenn die Korrelation zwischen Symptom und Diagnose negativ war, und er verhinderte, dass die Urteilenden Zusammenhänge erkannten, die tatsächlich vorhanden waren.

Die Verfügbarkeit bietet eine natürliche Erklärung für den Effekt der illusorischen Korrelation. Die Einschätzung der Häufigkeit des gemeinsamen Auftretens zweier Ereignisse könnte sich nach der Stärke der assoziativen Verknüpfung zwischen ihnen richten. Wenn die Verknüpfung stark ist, wird man vermutlich daraus folgern, dass die Ereignisse häufig paarweise aufgetreten sind. Nach dieser Auffassung ist beispielsweise die illusorische Korrelation zwischen Argwohn und einer eigentümlichen Zeichnung der Augen auf die Tatsache zurückzuführen, dass Argwohn unter allen Körperteilen am ehesten mit den Augen assoziiert wird.

Lebenslange Erfahrung hat uns gelehrt, dass im Allgemeinen Beispiele von großen Klassen besser und schneller abgerufen werden als Beispiele seltenerer Klassen; dass wahrscheinliche Ereignisse leichter vorstellbar sind als unwahrscheinliche; und dass die assoziativen Verknüpfungen zwischen Ereignissen verstärkt werden, wenn die Ereignisse häufig gemeinsam auftreten. Infolgedessen steht dem Menschen ein Verfahren (die Verfügbarkeitsheuristik) zur Verfügung, mit dem er anhand der Leichtigkeit, mit der die relevanten mentalen Operationen – Abruf, Konstruktion oder Assoziation – ausgeführt werden können, die Größe *(numerosity)* einer Klasse, die Wahrscheinlichkeit eines Ereignisses oder die Häufigkeit des gemeinsamen Auftretens abschätzen kann. Doch wie die vorhergehenden Beispiele gezeigt haben, führt dieses nützliche Schätzverfahren zu systematischen Fehlern.

## Anpassung und Verankerung

In vielen Situationen nehmen wir Schätzungen vor, indem wir bei einem Anfangswert beginnen, der so lange korrigiert wird, bis man die endgültige Lösung erhält. Der Anfangswert oder Ausgangspunkt mag durch die Formulierung des Problems nahegelegt werden oder das Ergebnis einer teilweisen Berechnung sein. In beiden Fällen sind die Korrekturen in der Regel unzureichend.[19] Das heißt, verschiedene Ausgangspunkte ergeben unterschiedliche Schätzungen, die in Richtung der Anfangswerte verzerrt sind. Wir nennen dieses Phänomen »Ankereffekt«.

*Unzureichende Anpassung.* In einer Demonstration des Ankereffekts sollten die Probanden verschiedene Größen (in Prozentsätzen) schätzen (zum Beispiel den Prozentsatz afrikanischer Länder in den Vereinten Nationen). Für jede Größe wurde durch Drehen eines Glücksrads in Anwesenheit der Versuchspersonen eine Zahl zwischen null und hundert festgelegt. Die Probanden sollten als Erstes angeben, ob diese Zahl höher oder niedriger war als der Wert der Größe, und anschließend den Wert der Größe abschätzen, indem sie sich von der gegebenen Zahl nach oben oder unten bewegten. Verschiedene Gruppen erhielten für jede Größe verschiedene Zahlen, und diese willkürlich ausgewählten Zahlen wirkten sich deutlich auf die Schätzungen aus. So betrugen beispielsweise die mittleren Schätzungen für den Prozentsatz afrikanischer Staaten in den Vereinten Nationen für Gruppen, die 10 beziehungsweise 65 als Ausgangspunkte erhalten hatten, 25 beziehungsweise 45. Belohnungen für Genauigkeit haben den Ankereffekt nicht vermindert.

Der Ankereffekt tritt nicht nur auf, wenn die Versuchsperson einen Ausgangswert erhält, sondern auch dann, wenn der Proband seine Schätzung auf das Ergebnis einer unvollständigen Berechnung stützt. Eine Studie über intuitive numerische Schätzungen verdeutlicht diesen Effekt. Zwei Gruppen von Highschool-Schülern sollten innerhalb von fünf Sekunden das Ergebnis eines numerischen Ausdrucks schätzen, der auf eine Tafel geschrieben wurde. Eine Gruppe schätzte das Produkt von

$$8 \times 7 \times 6 \times 5 \times 4 \times 3 \times 2 \times 1$$

während eine andere Gruppe folgendes Produkt schätzte:

$$1 \times 2 \times 3 \times 4 \times 5 \times 6 \times 7 \times 8$$

Um solche Fragen schnell zu beantworten, mögen Menschen einige Berechnungsschritte durchführen und das Produkt durch Extrapolation oder Anpassung abschätzen. Weil Anpassungen in der Regel unzureichend sind, sollte dieses Verfahren zu einer zu niedrigen Schätzung führen. Da das Ergebnis der ersten Multiplikationsschritte (von links nach rechts) in der fallenden Folge höher ist als in der steigenden Folge, sollte der erste Ausdruck als größer eingeschätzt werden als der zweite. Beide Vorhersagen wurden bestätigt. Die mittlere Schätzung für die steigende Folge war 512, während die mittlere Schätzung für die fallende Folge 2250 war. Die richtige Antwort lautet 40 320.

*Verzerrungen bei der Beurteilung konjunkter und disjunkter Ereignisse.* In einer neueren Studie von Bar- Hillel[20] hatten die Probanden die Gelegenheit, auf eines von zwei Ereignissen zu wetten. Drei Typen von Ereignissen wurden verwendet: (i) einfache Ereignisse wie etwa das Ziehen einer roten Murmel aus einem Beutel, der 50 Prozent rote und 50 Prozent weiße Murmeln enthält; (ii) konjunkte Ereignisse, etwa das Ziehen einer roten Murmel siebenmal hintereinander, mit Ersetzung, aus einem Beutel, der 90 Prozent rote Murmeln und 10 Prozent weiße Murmeln enthält; und (iii) disjunkte Ereignisse, wie etwa das Ziehen einer roten Murmel zumindest einmal in sieben aufeinanderfolgenden Versuchen, mit Ersetzung, aus einem Beutel, der 10 Prozent rote Murmeln und 90 Prozent weiße Murmeln enthält. Bei diesem Problem zog es eine deutliche Mehrheit der Probanden vor, auf das konjunkte Ereignis zu wetten (dessen Wahrscheinlichkeit 0,48 ist), statt das einfache Ereignis (dessen

Wahrscheinlichkeit 0,50 beträgt) vorzuziehen. Die Versuchspersonen wetteten außerdem eher auf das einfache als auf das disjunkte Ereignis, das eine Wahrscheinlichkeit von 0,52 hat. Folglich wetten die meisten Probanden in beiden Vergleichen auf das weniger wahrscheinliche Ereignis. Dieses Wahlmuster veranschaulicht einen allgemeinen Befund. Studien über Wahlen zwischen Lotterien und über Wahrscheinlichkeitsurteile deuten darauf hin, dass Menschen tendenziell die Wahrscheinlichkeit konjunkter Ereignisse zu hoch schätzen,[21] während sie die Wahrscheinlichkeit disjunkter Ereignisse zu niedrig ansetzen. Diese Verzerrungen lassen sich ohne Weiteres als Ankereffekte erklären. Die angegebene Wahrscheinlichkeit des Elementarereignisses (Erfolg in einem beliebigen Durchgang) stellt einen natürlichen Ausgangspunkt für die Schätzung der Wahrscheinlichkeiten sowohl konjunkter als auch disjunkter Ereignisse dar. Da die Anpassung ausgehend vom Ausgangspunkt typischerweise unzureichend ist, bleiben die endgültigen Schätzungen in beiden Fällen allzu nahe bei den Wahrscheinlichkeiten der Elementarereignisse. Man beachte, dass die Gesamtwahrscheinlichkeit eines konjunkten Ereignisses niedriger ist als die Wahrscheinlichkeit jedes Elementarereignisses, während die Gesamtwahrscheinlichkeit eines disjunkten Ereignisses höher ist als die Wahrscheinlichkeit jedes Elementarereignisses. Infolge der Ankerung wird die Gesamtwahrscheinlichkeit bei konjunkten Problemen zu hoch und bei disjunkten Problemen zu niedrig geschätzt.

Verzerrungen bei der Beurteilung verbundener Ereignisse sind im Rahmen von Planungsprozessen besonders bedeutsam. Der erfolgreiche Abschluss eines Projekts, wie etwa der Entwicklung eines neuen Produktes, hat üblicherweise einen konjunkten Charakter: Damit das Projekt zu einem Erfolg wird, muss jedes Ereignis aus einer Reihe von Ereignissen eintreten. Selbst wenn die Eintrittswahrscheinlichkeit jedes dieser Ereignisse sehr hoch ist, kann die Gesamtwahrscheinlichkeit eines erfolgreichen Abschlusses recht niedrig sein, wenn die Anzahl der Ereignisse groß ist. Die allgemeine Tendenz zur Überschätzung der Wahrscheinlichkeit konjunkter Ereignisse führt zu einem ungerechtfertigten Optimismus bei der Beurteilung der Wahrscheinlichkeit, dass ein Plan erfolgreich umgesetzt oder ein Projekt termingerecht zum Abschluss gebracht wird. Umgekehrt begegnet man disjunkten Strukturen typischerweise bei der Beurteilung von Risiken. In einem komplexen System wie etwa einem Kernreaktor oder einem menschlichen Körper kommt es zu einer Fehlfunktion, wenn eine

seiner wesentlichen Komponenten versagt. Selbst wenn die Wahrscheinlichkeit des Versagens der einzelnen Komponenten niedrig ist, kann die Wahrscheinlichkeit eines Gesamtausfalls hoch sein, wenn viele Komponenten betroffen sind. Wegen der Ankerung unterschätzen Menschen tendenziell die Wahrscheinlichkeiten des Versagens von komplexen Systemen. Daher lässt sich die Richtung der Ankerverzerrung manchmal aus der Struktur des Ereignisses ableiten. Die kettenartige Struktur von Konjunktionen führt zu Überschätzung, die trichterförmige Struktur von Disjunktionen führt zu Unterschätzung.

*Ankereffekte bei der Beurteilung subjektiver Wahrscheinlichkeitsverteilungen.* Bei der Entscheidungsanalyse sollen Experten oftmals ihre Einschätzung einer Größe, etwa des Wertes des Dow-Jones-Index an einem bestimmten Tag, in Form einer Wahrscheinlichkeitsverteilung ausdrücken. Um eine solche Verteilung zu konstruieren, wird die Person im Allgemeinen gebeten, Werte der Größe auszuwählen, die spezifischen Perzentilen ihrer subjektiven Wahrscheinlichkeitsverteilung entsprechen. So mag der Urteilende aufgefordert werden, eine Zahl, $X_{90}$, auszuwählen, und zwar so, dass seine subjektive Wahrscheinlichkeit, dass diese Zahl höher ist als der Wert des Dow-Jones-Index, 0,90 beträgt. Das heißt, er sollte den Wert $X_{90}$ so auswählen, dass er mit einer Wahrscheinlichkeit von neun zu eins davon ausgeht, dass der Dow-Jones-Index diesen nicht übertreffen wird. Eine subjektive Wahrscheinlichkeitsverteilung für den Wert des Dow-Jones-Index lässt sich aus mehreren solchen Urteilen konstruieren, die verschiedenen Perzentilen entsprechen.

Durch Sammlung subjektiver Wahrscheinlichkeitsverteilungen für viele verschiedene Größen kann man den Urteilenden auf seine ordnungsgemäße Kalibrierung testen. Ein Urteilender ist bei einer Reihe von Problemen ordnungsgemäß (oder extern) kalibriert, wenn genau $\Pi$ Prozent der wahren Werte der beurteilten Größen unter die von ihm angegebenen Werte von $X_{\Pi}$ fallen. So sollten zum Beispiel die wahren Werte für 1 Prozent der Größen unter $X_{01}$ und für 1 Prozent der Größen über $X_{99}$ fallen. Folglich sollten die wahren Werte bei 98 Prozent der Probleme in das Konfidenzintervall zwischen $X_{01}$ und $X_{99}$ fallen.

Mehrere Forscher[22] haben von einer Vielzahl von Urteilenden Wahrscheinlichkeitsverteilungen für viele Größen erhalten. Diese Verteilungen zeigten große und systematische Abweichungen von einer ordnungsgemäßen (»sauberen«) Kalibrierung. In den meisten Studien sind die tatsächlichen Werte der geschätzten Größen bei etwa 30 Prozent der Probleme entweder kleiner als $X_{01}$ oder größer als $X_{99}$. Das heißt, die Versuchsteilnehmer geben allzu enge Konfidenzintervalle an, in denen sich ein höheres Maß an Sicherheit widerspiegelt, als durch ihr Wissen über die geschätzten Größen gerechtfertigt ist. Diese Verzerrung ist sowohl bei Laien als auch bei Experten weit verbreitet, und sie lässt sich auch durch Einführung sachgerechter Bewertungsregeln, die Anreize für eine externe Kalibrierung liefern, nicht beseitigen. Dieser Effekt ist zumindest teilweise auf die Ankerung zurückzuführen.

Um zum Beispiel $X_{90}$ für den Wert des Dow-Jones-Index auszuwählen, liegt es nahe, mit der besten subjektiven Schätzung für den Dow-Jones-Index zu beginnen und diesen Wert anschließend nach oben zu korrigieren. Wenn diese Korrektur – wie die meisten anderen – unzureichend ist, wird $X_{90}$ nicht hinreichend extrem sein. Ein ähnlicher Ankereffekt kommt bei der Auswahl von $X_{10}$ zum Tragen, den man wahrscheinlich dadurch erhält, dass man seine beste Schätzung nach unten korrigiert. Entsprechend wird das Konfidenzintervall zwischen $X_{10}$ und $X_{90}$ zu eng sein, und die geschätzte Wahrscheinlichkeitsverteilung wird zu schmal sein. Für diese Interpretation spricht auch, dass die subjektiven Wahrscheinlichkeiten durch ein Verfahren, bei dem die beste Schätzung nicht als Anker dient, systematisch verändert werden können.

Es gibt zwei verschiedene Methoden, um subjektive Wahrscheinlichkeitsverteilungen für eine gegebene Größe (den Dow-Jones-Index) zu erhalten: (i) indem man den Versuchsteilnehmer bittet, Werte für den Dow-Jones-Index auszuwählen, die genau angegebenen Perzentilen seiner Wahrscheinlichkeitsverteilung entsprechen, und (ii) indem man den Probanden bittet, die Wahrscheinlichkeiten zu schätzen, dass der wahre Wert des Dow-Jones-Index einige genau angegebene Werte übertreffen wird. Die beiden Verfahren sind formal äquivalent und sollten gleiche Verteilungen erbringen. Allerdings legen sie verschiedene Modi der Anpassung von unterschiedlichen Ankern nahe. Beim Verfahren (i) ist der natürliche Ausgangspunkt die beste subjektive Schätzung des Befragten.

Bei Verfahren (ii) andererseits ist der Proband möglicherweise in dem in der Frage angegebenen Wert »verankert«. Wahlweise mag er in der Gleichwahrscheinlichkeit beziehungsweise einer 50-Prozent-Chance verankert sein, die ein natürlicher Ausgangspunkt bei der Wahrscheinlichkeitsschätzung ist. In beiden Fällen sollte Verfahren (ii) weniger extreme Wahrscheinlichkeiten ergeben als Verfahren (i).

Um die beiden Verfahren einander gegenüberzustellen, wurde eine Menge von 24 Größen (wie etwa die Luftlinie von New Delhi nach Peking) einer Gruppe von Probanden präsentiert, die entweder $X_{10}$ oder $X_{90}$ für jedes Problem schätzen sollte. Einer anderen Probandengruppe wurde der Mittelwert der Schätzungen der ersten Gruppe für jede der 24 Größen mitgeteilt. Sie wurden gebeten, die Wahrscheinlichkeit zu beurteilen, dass jeder der gegebenen Werte den wahren Wert der relevanten Größe übertrifft. Wenn keine Verzerrung vorliegt, sollte die zweite Gruppe auf die Wahrscheinlichkeit zurückgreifen, die der ersten Gruppe mitgeteilt wurde, nämlich 9 : 1. Doch wenn die Gleichwahrscheinlichkeit oder der angegebene Wert als Anker dienen, sollte die geschätzte Wahrscheinlichkeit der zweiten Gruppe weniger extrem sein, das heißt näher bei 1 : 1 liegen. Tatsächlich betrug die von dieser Gruppe angegebene mittlere Wahrscheinlichkeit, über alle Probleme hinweg, 3 : 1. Als die Schätzungen beider Gruppen auf externe Kalibrierung getestet wurden, fand man heraus, dass die Probanden in der ersten Gruppe zu extrem waren, was mit den Ergebnissen früherer Studien übereinstimmte. Die Ereignisse, denen sie eine Wahrscheinlichkeit von 0,10 zuschrieben, traten in 24 Prozent der Fälle ein. Dagegen waren die Probanden in der zweiten Gruppe zu vorsichtig. Ereignisse, denen sie eine mittlere Wahrscheinlichkeit von 0,34 zuschrieben, traten in 26 Prozent der Fälle ein. Diese Ergebnisse veranschaulichen das Ausmaß, in dem der Kalibrierungsgrad vom Erhebungsverfahren abhängig ist.

## Diskussion

Dieser Aufsatz befasste sich mit kognitiven Verzerrungen, die auf die Anwendung von Urteilsheuristiken zurückzuführen sind. Diese Verzerrungen sind nicht motivationalen Effekten wie Wunschdenken oder der Verzerrung von Urteilen durch Belohnungen und Bestrafungen zuzuschreiben. Tatsächlich ereigneten sich etliche der weiter vorn beschriebenen gravierenden Urteilsfehler ungeachtet der Tatsache, dass die Versuchsteilnehmer ermuntert wurden, präzise Angaben zu machen, und für die richtigen Antworten belohnt wurden.[23]

Der Rückgriff auf Heuristiken und die weite Verbreitung von Verzerrungen sind nicht auf Laien beschränkt. Erfahrene Forscher sind ebenfalls anfällig für die gleichen Verzerrungen – wenn sie intuitiv denken. So wurde zum Beispiel die Neigung, das Ergebnis, das die Daten am besten repräsentiert, ohne hinlängliche Berücksichtigung der A-priori-Wahrscheinlichkeit vorherzusagen, bei den intuitiven Urteilen von Personen beobachtet, die eine gründliche Schulung in Statistik erhalten hatten.[24] Obgleich Personen mit Statistik-Kenntnissen elementare Fehler wie den Spielerfehlschluss vermeiden, unterliegen ihre intuitiven Urteile bei verwickelteren und weniger transparenten Problemen ähnlichen Fehlschlüssen.

Es ist nicht weiter verwunderlich, dass nützliche Heuristiken wie Repräsentativität und Verfügbarkeit beibehalten werden, auch wenn sie gelegentlich zu Vorhersage- und Schätzfehlern führen. Verwunderlich ist vielleicht der Umstand, dass Menschen aus lebenslangen Erfahrungen nicht fundamentale statistische Regeln wie die Regression zum Mittelwert oder auch die Auswirkung des Stichprobenumfangs auf die Stichprobenvarianz ableiten. Obgleich jeder Mensch im normalen Verlauf seines Lebens zahlreichen Beispielen ausgesetzt ist, aus denen diese Regeln abgeleitet werden könnten, entdecken nur sehr wenige von sich aus die Grundsätze statistischer Auswahlverfahren und der Regression. Die Alltagserfahrung lehrt uns keine statistischen Prinzipien, weil die relevanten Beispiele nicht in geeigneter Weise kodiert sind. So bemerken wir zum Beispiel nicht, dass die durchschnittliche Wortlänge in aufeinanderfolgenden Textzeilen stärker variiert als auf aufeinanderfolgenden Seiten, weil wir der durchschnittlichen

Wortlänge in einzelnen Zeilen oder Seiten keine Beachtung schenken. Daher lernen Menschen nicht den Zusammenhang zwischen Stichprobenumfang und Stichprobenvarianz, obgleich es reichlich Daten gäbe, um diesen zu erkennen.

Das Fehlen eines angemessenen Codes erklärt auch, weshalb Menschen für gewöhnlich die Verzerrungen in ihren Wahrscheinlichkeitsurteilen nicht bemerken. Eine Person könnte möglicherweise herausfinden, ob ihre Urteile extern kalibriert sind, wenn sie über den Prozentsatz der Ereignisse Buch führen würde, die unter denjenigen, denen sie die gleiche Wahrscheinlichkeit zuschreibt, tatsächlich eintreten. Allerdings klassifizieren wir Ereignisse nicht automatisch nach ihrer geschätzten Wahrscheinlichkeit. Ohne eine solche Klassifizierung ist es für ein Individuum zum Beispiel unmöglich, herauszufinden, dass sich nur 50 Prozent der Vorhersagen, denen es eine Wahrscheinlichkeit von 0,90 oder höher zuschrieb, bewahrheitet haben.

Die empirische Analyse kognitiver Verzerrungen hat Konsequenzen für die theoretische und praktische Rolle geschätzter Wahrscheinlichkeiten. Die moderne Entscheidungstheorie[25] betrachtet die subjektive Wahrscheinlichkeit als die quantifizierte Meinung einer idealisierten Person. Insbesondere ist die subjektive Wahrscheinlichkeit eines gegebenen Ereignisses definiert durch die Menge der Wetten über dieses Ereignisses, die eine solche Person einzugehen bereit wäre. Ein in sich konsistentes oder kohärentes subjektives Wahrscheinlichkeitsmaß lässt sich dann für ein Individuum ableiten, wenn seine Wahlen zwischen Wetten gewissen Grundsätzen genügen, nämlich den Axiomen der Theorie. Die abgeleitete Wahrscheinlichkeit ist in dem Sinne subjektiv, als verschiedene Individuen die Wahrscheinlichkeit desselben Ereignisses unterschiedlich einschätzen können. Die hauptsächliche Stärke dieses Ansatzes besteht darin, dass er eine strenge subjektive Interpretation der Wahrscheinlichkeit liefert, die auf einzigartige Ereignisse anwendbar und in eine allgemeine Theorie der rationalen Entscheidung eingebettet ist.

Vielleicht sollte angemerkt werden, dass sich subjektive Wahrscheinlichkeiten zwar manchmal aus Präferenzen auf Wetten ableiten lassen, im Allgemeinen aber nicht auf diese Weise gebildet werden. Eine Person setzt auf Team A statt auf Team B, weil sie glaubt, dass Team A bessere Siegeschancen hat; sie leitet diese Überzeugung nicht aus ihren Wettpräferenzen ab. In Wirklichkeit bestimmen daher die subjektiven

Wahrscheinlichkeiten die Präferenzen zwischen Wetten und werden nicht aus diesen abgeleitet, wie es in der axiomatischen Theorie rationaler Entscheidungen angenommen wird.[26]

Die an sich subjektive Natur der Wahrscheinlichkeit hat viele Wissenschaftler zu der Überzeugung gebracht, dass Kohärenz oder innere Widerspruchsfreiheit das einzige gültige Kriterium sei, nach dem geschätzte Wahrscheinlichkeiten beurteilt werden sollten. Vom Standpunkt der formalen Theorie subjektiver Wahrscheinlichkeit ist jede Menge in sich widerspruchsfreier Wahrscheinlichkeitsurteile so gut wie jede andere. Dieses Kriterium ist nicht völlig zufriedenstellend, weil eine in sich widerspruchsfreie Menge subjektiver Wahrscheinlichkeiten mit anderen Überzeugungen des Individuums unvereinbar sein kann. Betrachten wir eine Person, in deren subjektiven Wahrscheinlichkeiten sich für alle möglichen Ergebnisse eines Münzwurfspiels der Spielerfehlschluss widerspiegelt. Das heißt, ihre Schätzung für die Wahrscheinlichkeit von »Zahl« bei einem bestimmten Wurf steigt mit der Anzahl aufeinanderfolgender »Kopf«-Ergebnisse, die diesem Wurf vorangingen. Die Urteile einer solchen Person könnten in sich widerspruchsfrei und daher gemäß dem Kriterium der formalen Theorie als adäquate subjektive Wahrscheinlichkeiten akzeptabel sein. Diese Wahrscheinlichkeiten sind jedoch nicht mit der allgemein anerkannten Überzeugung vereinbar, dass eine Münze kein Gedächtnis hat und daher keine sequenziellen Abhängigkeiten erzeugen kann. Damit geschätzte Wahrscheinlichkeiten als angemessen oder rational beurteilt werden können, reicht innere Widerspruchsfreiheit nicht aus. Die Urteile müssen mit dem gesamten Netz von Überzeugungen eines Individuums vereinbar sein. Leider gibt es kein einfaches formales Verfahren, um zu beurteilen, ob eine Serie von Wahrscheinlichkeitsurteilen mit dem gesamten System von Überzeugungen des Urteilenden vereinbar ist. Der rational Urteilende wird dennoch nach Vereinbarkeit streben, auch wenn die innere Widerspruchsfreiheit leichter zu bewerkstelligen und zu beurteilen ist. Insbesondere wird er versuchen, seine Wahrscheinlichkeitsurteile mit seinem Wissen über das Thema, den Gesetzen der Wahrscheinlichkeit und seinen eigenen Urteilsheuristiken und kognitiven Verzerrungen in Einklang zu bringen.

## Zusammenfassung

Dieser Artikel beschrieb drei Heuristiken, die bei der Urteilsbildung unter Unsicherheit benutzt werden: (i) Repräsentativität, die für gewöhnlich angewendet wird, wenn Menschen die Wahrscheinlichkeit schätzen sollen, mit der ein Objekt oder Ereignis A zur Klasse oder zum Prozess B gehört; (ii) Verfügbarkeit von Beispielen oder Szenarien, die oftmals benutzt wird, wenn Menschen die Häufigkeit einer Klasse oder die Plausibilität einer bestimmten Entwicklung beurteilen sollen; und (iii) Anpassung an einen Anker, die für gewöhnlich bei numerischen Vorhersagen angewandt wird, wenn ein relevanter Wert verfügbar ist. Diese Heuristiken sind höchst ökonomisch und für gewöhnlich effektiv, aber sie führen zu systematischen und vorhersagbaren Irrtümern. Ein besseres Verständnis dieser Heuristiken und der Verzerrungen, zu denen sie führen, könnte Urteile und Entscheidungen in Situationen der Unsicherheit verbessern.

# *Entscheidungen, Werte und Frames*[1]

Daniel Kahneman und Amos Tversky

Abriss: Wir diskutieren die kognitiven und psychophysischen Determinanten von Wahlen in riskanten und risikolosen Kontexten. Die Psychophysik des Wertes erzeugt Risikoscheu im Bereich von Gewinnen und Risikofreude im Bereich von Verlusten. Die Psychophysik der Wahrscheinlichkeit erzeugt eine Übergewichtung sicherer Optionen und unwahrscheinlicher Ereignisse im Vergleich zu Ereignissen mittlerer Wahrscheinlichkeit. Entscheidungsprobleme können auf vielfältige Weise beschrieben beziehungsweise »gerahmt« *(framed)* werden und so verschiedene Präferenzen erzeugen, was dem Invarianzkriterium der rationalen Entscheidungsfindung widerspricht. Der Prozess der mentalen Buchführung, in dem Menschen die Ergebnisse von Transaktionen organisieren, erklärt einige Anomalien des Verbraucherverhaltens. Insbesondere kann die Annehmbarkeit einer Option davon abhängen, ob ein negatives Ergebnis als ein Kostenfaktor oder ein nicht ersetzter Verlust bewertet wird. Die Beziehung zwischen Entscheidungswerten und Erfahrungswerten wird diskutiert.

Menschen treffen Entscheidungen mit der gleichen Selbstverständlichkeit, mit der sie reden – sie tun es wissentlich oder unwissentlich in einem fort. Daher ist es nicht weiter verwunderlich, dass sich viele Disziplinen, von der Mathematik und Statistik über die Wirtschafts- und Politikwissenschaften bis zur Soziologie und Psychologie, mit dem Thema »Entscheidungsfindung« befassen. Die Entscheidungsforschung behandelt sowohl normative als auch deskriptive Fragestellungen. Die normative Analyse befasst sich mit der Rationalität und Logik der Entscheidungsfindung, die deskriptive Analyse dagegen mit den Überzeugungen und Präferenzen von Menschen, so wie sie sind, nicht so, wie sie sein sollten. Das Spannungsverhältnis zwischen normativen und deskriptiven Betrachtungen kennzeichnet einen Großteil der Studien über Urteils- und Entscheidungsfindung.

Analysen der Entscheidungsfindung unterscheiden für gewöhnlich zwischen riskanten und risikolosen Entscheidungen. Das paradigmatische Beispiel für eine Entscheidung unter Risiko ist das Akzeptieren einer Lotterie, die monetäre Ergebnisse mit eindeutigen Wahrscheinlichkeiten zeitigt. Eine typische risikolose Entscheidung betrifft die Annehmbarkeit einer Transaktion, bei der ein Gut oder eine Dienstleistung gegen Geld oder Arbeit getauscht wird. Im ersten Teil dieses Beitrags präsentieren wir eine Analyse der kognitiven und psychophysischen Faktoren, die den Wert riskanter Gewinnaussichten *(prospects)* bestimmen. Im zweiten Teil erweitern wir diese Analyse auf Transaktionen und Tauschprozesse *(trades)*.

## Risikobehaftete Entscheidungen

Risikoentscheidungen, wie etwa die Frage, ob man einen Regenschirm mitnehmen oder einen Krieg anfangen soll, werden getroffen, ohne dass man im Vorhinein ihre Folgen absehen könnte. Weil die Konsequenzen solcher Handlungen von ungewissen Ereignissen wie etwa dem Wetter oder der Entschlossenheit des Feindes abhängen, kann man die Wahl einer Handlungsoption als die Annahme einer Lotterie interpretieren, bei der verschiedene Ergebnisse mit unterschiedlichen Wahrscheinlichkeiten eintreten können. Es ist daher verständlich, dass sich die Erforschung von Entscheidungsprozessen unter Risiko auf Wahlen zwischen einfachen Lotterien mit monetären Ergebnissen und eindeutigen Wahrscheinlichkeiten konzentriert hat, in der Hoffnung, dass diese einfachen Probleme grundlegende Einstellungen zu Risiko und [subjektivem] Wert enthüllen.

Wir skizzieren ein Rahmenmodell für die Betrachtung von Risikoentscheidungen, das viele seiner Hypothesen aus einer psychophysischen Analyse von Reaktionen auf Geld und Wahrscheinlichkeit ableitet. Die psychophysische Herangehensweise an die Entscheidungsfindung geht zurück auf einen bemerkenswerten Aufsatz, den Daniel Bernoulli 1738 veröffentlichte.[2] Darin versuchte er zu erklären, wieso Menschen im Allgemeinen risikoscheu sind und weshalb die Risikoscheu mit zunehmendem Wohlstand abnimmt. Betrachten wir, zur Veranschaulichung des Phänomens der Risikoscheu und von Bernoullis Analyse, die Wahl zwischen der Aussicht, mit einer 85-prozentigen Wahrscheinlichkeit 1000 Dollar zu bekommen (und einer 15-prozentigen Wahrscheinlichkeit, nichts zu gewinnen), und der Alternative, 800 Dollar sicher zu erhalten. Eine große Mehrheit der Befragten zieht die sichere Option der Lotterie vor, obwohl die Lotterie einen höheren (mathematischen) Erwartungswert hat. Der Erwartungswert einer Lotterie ist ein gewichteter Mittelwert, bei dem jedes mögliche Ergebnis mit seiner Eintrittswahrscheinlichkeit gewichtet wird. Der Erwartungswert der Lotterie in diesem Beispiel beträgt 0,85 × 1000 Dollar + 0,15 × 0 Dollar = 850 Dollar, was über dem Erwartungswert der sicheren Alternative (800 Dollar) liegt. Die Präferenz für den sicheren Gewinn ist ein Beispiel für Risikoscheu. Eine Präferenz für ein sicheres Ergebnis gegenüber einer

Lotterie mit einem höheren oder gleichen Erwartungswert wird im Allgemeinen als »Risikoaversion« bezeichnet, während die Ablehnung einer sicheren Option zugunsten einer Lotterie mit niedrigerem oder gleichem Erwartungswert »Risikofreude« genannt wird.

Bernoulli behauptete, Menschen würden Gewinnaussichten *(prospects)* nicht nach dem Erwartungswert ihrer monetären Ergebnisse, sondern nach dem Erwartungswert des subjektiven Werts dieser Ergebnisse beurteilen. Der subjektive Wert einer Lotterie seinerseits ist ebenfalls ein gewichteter Mittelwert, aber jetzt wird der subjektive Wert jedes Ergebnisses mit dessen Wahrscheinlichkeit gewichtet. Um Risikoscheu innerhalb dieses Rahmenmodells erklären zu können, machte Bernoulli die Annahme, der subjektive Wert oder Nutzen sei eine konkave Funktion des Geldes. In einer solchen Funktion ist die Differenz zwischen den Nutzwerten von 200 und 100 Dollar zum Beispiel größer als der Nutzwertunterschied zwischen 1200 und 1100 Dollar. Aus der Konkavität folgt, dass der subjektive Wert, der einem Gewinn von 800 Dollar zugeschrieben wird, mehr als 80 Prozent des Wertes eines Gewinns von 1000 Dollar ausmacht. Aus der Konkavität der Nutzenfunktion folgt daher eine risikoaverse Präferenz für einen sicheren Gewinn von 800 Dollar gegenüber einer 80-prozentigen Chance auf einen Gewinn von 1000 Dollar, obgleich die beiden Aussichten denselben monetären Erwartungswert besitzen.

In der Entscheidungsanalyse ist es üblich, die Entscheidungsergebnisse nach der Gesamtvermögenshöhe zu beschreiben. So wird zum Beispiel das Angebot, 20 Dollar auf einen fairen Münzwurf zu setzen, als eine Wahl zwischen der gegenwärtigen Vermögensposition W eines Individuums und einer gleich hohen Wahrscheinlichkeit dargestellt, sich zu W + 20 Dollar oder W – 20 Dollar zu bewegen. Diese Darstellungsweise erscheint psychologisch unrealistisch: Menschen stellen sich vergleichsweise unbedeutende Entscheidungsergebnisse nicht als Vermögenszustände vor, sondern als Gewinne, Verluste und neutrale Ergebnisse (wie zum Beispiel die Beibehaltung des Status quo). Wenn Vermögensänderungen und nicht endgültige Vermögenszustände die tatsächlichen Träger subjektiver Nutzwerte sind, wie wir behaupten, dann sollte die psychophysische Analyse der Ergebnisse auf Gewinne und Verluste und nicht auf das Gesamtvermögen angewandt werden. Diese Annahme spielt eine zentrale Rolle in einem Modell von Risikoentscheidungen, das wir »Neue Erwartungstheorie« *(Prospect Theory)*[3] genannt haben. Introspektion und

psychophysische Messungen deuten darauf hin, dass der subjektive Wert eine konkave Funktion der Höhe eines Gewinns ist. Die gleiche Verallgemeinerung gilt auch für Verluste. Die Differenz im subjektiven Wert zwischen einem Verlust von 200 Dollar und einem Verlust von 100 Dollar erscheint größer als die Differenz im subjektiven Wert zwischen einem Verlust von 1200 Dollar und einem Verlust von 1100 Dollar. Wenn die Wertfunktionen für Gewinne und für Verluste zusammengesetzt werden, erhalten wir eine s-förmige Funktion des in Abbildung 1 gezeigten Typs.

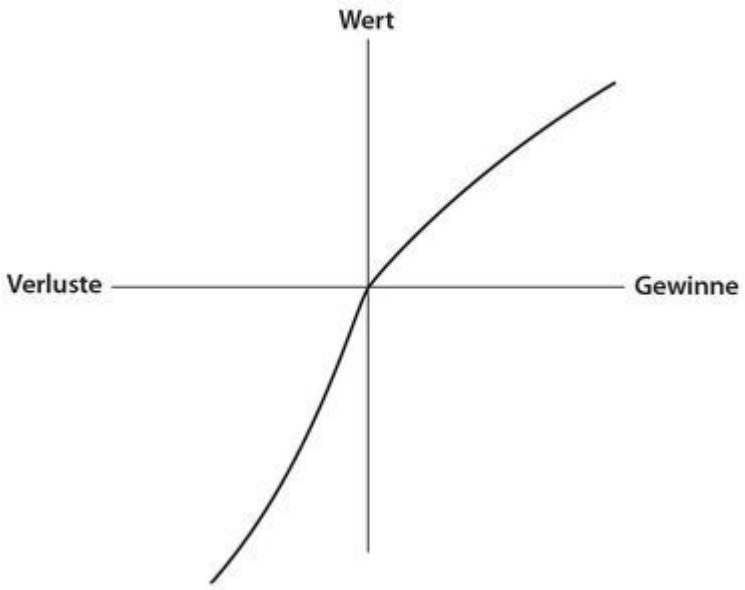


**Abbildung 1**
**Eine hypothetische Wertfunktion**

Die in Abbildung 1 gezeigte Wertfunktion ist (a) über Gewinne und Verluste und nicht über das Gesamtvermögen definiert, (b) im Bereich von Gewinnen konkav und im Bereich von Verlusten konvex und (c) für Verluste deutlich steiler als für Gewinne. Die letzte Eigenschaft, die wir »Verlustaversion« nennen, drückt die Intuition aus, dass die Abneigung gegen einen Verlust von X Dollar stärker ist als Vorfreude auf einen Gewinn von X Dollar. Die Verlustaversion erklärt, weshalb Menschen nur widerwillig auf einen fairen Münzwurf für gleiche Einsätze wetten: Die Attraktivität des möglichen Gewinns ist nicht annähernd ausreichend, um die Abneigung gegen den möglichen Verlust aufzuwiegen. So wollten die meisten Befragten in einer Stichprobe von Studenten keine 10 Dollar auf

den Wurf einer Münze setzen, wenn sie weniger als 30 Dollar gewinnen konnten.

Die Annahme der Risikoaversion spielt in der volkswirtschaftlichen Theorie eine zentrale Rolle. Doch genauso wie die Konkavität der Wertfunktion im Gewinnbereich eine Risikoaversion impliziert, geht mit der Konvexität der Wertfunktion im Verlustbereich Risikofreude einher. Tatsächlich ist Risikofreude bei Verlusten ein robuster Effekt, insbesondere wenn die Verlustwahrscheinlichkeiten erheblich sind. Betrachten wir zum Beispiel eine Situation, in der ein Individuum zwischen einer 85-prozentigen Chance, 1000 Dollar zu verlieren (und einer 15-prozentigen Chance, nichts zu verlieren), und einem sicheren Verlust von 800 Dollar wählen muss. Eine große Mehrheit der Befragten äußert eine Präferenz für die Lotterie gegenüber der sicheren Alternative. Dies ist eine risikofreudige Wahl, weil der Erwartungswert der Lotterie (minus 850 Dollar) dem Erwartungswert des sicheren Verlustes (minus 800 Dollar) unterlegen ist. Risikofreude im Verlustbereich wurde von mehreren Forschern bestätigt.[4] Sie wurde auch bei nicht monetären Ergebnissen wie Stunden von Schmerzerleben[5] und Verlusten an Menschenleben[6] beobachtet. Ist es falsch, risikoscheu im Gewinnbereich und risikofreudig im Verlustbereich zu sein? Diese Präferenzen entsprechen überzeugenden Intuitionen über die subjektive Wertschätzung von Gewinnen und Verlusten, und es wird angenommen, dass Menschen einen Anspruch auf ihre eigene Wertschätzung haben sollten. Wir werden jedoch sehen, dass eine s-förmige Wertfunktion normativ unannehmbare Folgen hat.

Zur Darstellung der normativen Problematik wenden wir uns von der Psychologie ab und der Entscheidungstheorie zu. Die moderne Entscheidungstheorie begann mit den bahnbrechenden Arbeiten von Neumann und Morgenstern [7] (1947), die mehrere qualitative Grundsätze oder Axiome aufstellten, denen die Präferenzen eines rationalen Entscheiders gehorchen sollen. Zu diesen Axiomen gehörten Transitivität (wenn A gegenüber B vorgezogen wird und B gegenüber C, dann wird A gegenüber C vorgezogen) und Substitution (wenn A gegenüber B vorgezogen wird, dann wird eine 50-Prozent-Chance, A oder C zu erhalten, einer 50-Prozent-Chance, B oder C zu erhalten, vorgezogen), zusammen mit anderen Bedingungen formalerer Natur. Der normative und der deskriptive Status der Axiome rationaler Wahl wurde ausgiebig diskutiert.

So gibt es insbesondere überzeugende Belege dafür, dass Menschen nicht immer dem Substitutionsaxiom gehorchen, und der normative Wert dieses Axioms ist in hohem Maße strittig.[8] Sämtliche Analysen rationaler Entscheidung beziehen allerdings zwei Prinzipien ein: Dominanz und Invarianz. Dominanz verlangt, dass, wenn Gewinnaussicht A in jeder Hinsicht mindestens genauso gut wie Gewinnaussicht B und in wenigstens einer Hinsicht besser als B ist, A gegenüber B vorgezogen werden sollte. Invarianz verlangt, dass die Präferenzordnung zwischen Gewinnaussichten nicht von der Art und Weise ihrer Beschreibung abhängig sein sollte. Insbesondere sollten zwei Versionen eines Wahlproblems, die bei gemeinsamer Darbietung als äquivalent erkannt werden, auch bei getrennter Darbietung die gleiche Präferenz auslösen. Wir zeigen jetzt, dass das Erfordernis der Invarianz, wie elementar und harmlos es auch erscheinen mag, nicht allgemein erfüllt werden kann.

## Framing der Ergebnisse

Riskante Aussichten sind durch ihre möglichen Ergebnisse und die Eintrittswahrscheinlichkeiten dieser Ergebnisse charakterisiert. Dieselbe Option lässt sich allerdings in unterschiedlicher Weise darstellen *(framen,* auch: »einrahmen«) oder beschreiben.[9] So lassen sich beispielsweise die möglichen Resultate einer Lotterie entweder als Gewinne und Verluste in Bezug auf den Status quo darstellen oder als Vermögenspositionen, die den anfänglichen Vermögenszustand einbeziehen. Die Invarianz verlangt, dass solche Veränderungen in der Beschreibung von Ergebnissen die Präferenzordnung nicht beeinflussen sollten. Das folgende Paar von Problemen veranschaulicht einen Verstoß gegen dieses Erfordernis. Die Gesamtzahl der Befragten bei jedem Problem wird mit N bezeichnet, und der Prozentsatz derjenigen, die sich für die jeweilige Option entscheiden, wird in Klammern angegeben.

> Problem 1 (*N* = 152): Stellen Sie sich vor, die USA bereiten sich auf den Ausbruch einer ungewöhnlichen asiatischen Krankheit vor, die erwartungsgemäß 600 Menschenleben fordern soll. Zwei alternative Pläne zur Bekämpfung der Krankheit wurden vorgeschlagen. Angenommen, die exakten wissenschaftlichen Schätzungen der Konsequenzen der Pläne lauten folgendermaßen:
>
> > Wird Plan A umgesetzt, werden 200 Menschenleben gerettet. (72 Prozent)
> >
> > Wird Plan B umgesetzt, besteht eine Wahrscheinlichkeit von einem Drittel, dass 600 Menschenleben gerettet werden, und eine Wahrscheinlichkeit von zwei Dritteln, dass niemand gerettet wird. (28 Prozent)

Welchen der beiden Pläne würden Sie vorziehen?

Die Formulierung von Problem 1 nimmt stillschweigend einen Zustand als Referenzpunkt an, in dem die Krankheit 600 Menschenleben fordern darf. Die Ergebnisse der Programme schließen den Referenzzustand und zwei mögliche Gewinne ein, gemessen an der Anzahl der geretteten Menschenleben. Erwartungsgemäß sind die Präferenzen risikoavers: Eine klare Mehrheit der Befragten zieht es vor, 200 Menschen sicher zu retten, statt sich auf eine Lotterie einzulassen, die mit einer Chance von einem Drittel verbunden ist, 600 Menschenleben zu retten. Betrachten Sie jetzt ein anderes Problem, bei dem auf die gleiche Ausgangsgeschichte eine andere Beschreibung der mit den beiden Plänen verbundenen Aussichten folgt:

Problem 2 ($N$ = 155):

> Wird Plan C umgesetzt, werden 400 Menschen sterben. (27 Prozent) Wird Plan D umgesetzt, besteht eine Wahrscheinlichkeit von einem Drittel, dass niemand sterben wird, und eine Wahrscheinlichkeit von zwei Dritteln, dass 600 Menschen sterben werden. (78 Prozent)

Es lässt sich leicht nachweisen, dass zwischen den Optionen C und D bei Problem 2 und den Optionen A und B bei Problem 1 de facto kein Unterschied besteht. Allerdings geht die zweite Version von einem Referenzzustand aus, in dem niemand an der Krankheit stirbt. Das beste Ergebnis ist die Aufrechterhaltung dieses Zustands, und die Alternativen sind Verluste, gemessen an der Zahl der Menschen, die an der Krankheit sterben werden. Menschen, die Optionen in dieser Weise bewerten, sollten eine risikogeneigte Präferenz für die Lotterie (Option D) gegenüber dem sicheren Verlust von 400 Menschenleben zeigen. Tatsächlich ist die

Risikofreude in der zweiten Version des Problems größer als die Risikoaversion in der ersten Fassung.

Verstöße gegen das Invarianzaxiom sind weit verbreitet und robust. Sie sind sowohl unter fachkundigen wie unter naiven Befragten häufig anzutreffen, und sie lassen sich auch dann nicht beseitigen, wenn dieselben Personen beide Fragen innerhalb weniger Minuten beantworten. Konfrontiert man die Befragten mit ihren widersprüchlichen Antworten, sind sie typischerweise verblüfft. Selbst nach einem erneuten Durchlesen der Problemstellungen möchten sie weiterhin risikoavers in der Version »gerettete Menschenleben« sein; sie wollen in der Version »Verluste an Menschenleben« risikogeneigt sein. Und sie wollen auch weiterhin in beiden Versionen dem Invarianzprinzip genügen und konsistente Antworten geben. In ihrer hartnäckigen Anziehungskraft besitzen Framing-Effekte eine größere Ähnlichkeit mit Wahrnehmungstäuschungen als mit Rechenfehlern.

Das folgende Paar von Problemen ruft Präferenzen hervor, die die Dominanzbedingung rationaler Entscheidungsfindung verletzen.

> Problem 3 ($N$ = 86): Wählen Sie zwischen:
>
> E 25-prozentige Chance, 240 Dollar zu gewinnen, und 75-prozentige Chance, 760 Dollar zu verlieren. (0 Prozent)
> F 25-prozentige Chance, 250 Dollar zu gewinnen, und 75-prozentige Chance, 750 Dollar zu verlieren. (100 Prozent)

Es ist leicht ersichtlich, dass F gegenüber E dominiert. Tatsächlich wählen alle Befragten entsprechend.

> Problem 4 ($N$ = 150): Stellen Sie sich vor, Sie sind mit dem folgenden Paar konkurrierender Entscheidungen konfrontiert. Prüfen Sie zunächst beide Entscheidungen, und geben Sie dann die Optionen an, die Sie vorziehen.

Entscheidung 1: Wählen Sie zwischen:

A Sicherer Gewinn von 240 Dollar. (84 Prozent)
B 25-prozentige Chance, 1000 Dollar zu gewinnen, und 75-prozentige Chance, nichts zu gewinnen. (16 Prozent)

Entscheidung 2: Wählen Sie zwischen:

C Sicherer Verlust von 750 Dollar. (13 Prozent)
D 75-prozentige Chance, 1000 Dollar zu verlieren, und 25-prozentige Chance, nichts zu verlieren. (87 Prozent)

Wie nach der vorhergehenden Analyse zu erwarten, entschied sich bei der ersten Wahl eine große Mehrheit der Versuchsteilnehmer für den sicheren Gewinn gegenüber der positiven Lotterie, und eine noch größere Mehrheit der Probanden zeigte bei der zweiten Wahl eine risikogeneigte Präferenz für die Lotterie gegenüber dem sicheren Verlust. Tatsächlich entschieden sich 73 Prozent der Befragten für A und D und nur 3 Prozent für B und C. In einer modifizierten Version des Problems, mit geringeren Einsätzen, bei dem Studenten Lotterien auswählten, die sie tatsächlich spielen wollten, zeigte sich das gleiche Ergebnismuster.

Weil die Versuchspersonen die beiden Entscheidungen bei Problem 4 gleichzeitig erwogen, bekundeten sie in der Tat eine Präferenz für A und D gegenüber B und C. Doch wird die bevorzugte Verbindung faktisch von der abgelehnten dominiert. Addiert man den sicheren Gewinn von 240 Dollar (Option A) zu Option D, erhält man eine 25-prozentige Chance, 240 Dollar zu gewinnen, und eine 75-prozentige Chance, 760 Dollar zu verlieren. Genau dies ist Option F bei Problem 3. Somit führen die Anfälligkeit für Framing-Effekte und die s-förmige Wertfunktion bei einer Serie konkurrierender Entscheidungen zu einer Verletzung der Dominanzbedingung.

Die »Moral« dieser Ergebnisse ist verstörend: Invarianz ist normativ unabdingbar, intuitiv überzeugend und psychologisch unrealistisch. Tatsächlich können wir uns nur zwei Möglichkeiten vorstellen, um die Invarianz zu gewährleisten. Die erste besteht darin, eine Vorgehensweise zu wählen, die äquivalente Versionen jedes beliebigen Problems in die gleiche kanonische Repräsentation transformiert. Dies ist der Grund für die gängige

Ermahnung an Studenten der Betriebswirtschaftslehre, jedes Entscheidungsproblem unter dem Gesichtspunkt der Gesamtvermögenshöhe und nicht nach Gewinnen oder Verlusten zu erwägen.[10] Eine solche Repräsentation würde die in den vorangehenden Problemen veranschaulichten Verletzungen der Invarianz vermeiden, aber dieser Rat lässt sich leichter geben als befolgen. Außer in Situationen eines möglichen Ruins erscheint es selbstverständlicher, finanzielle Ergebnisse als Gewinne und Verluste statt als Vermögenszustände zu betrachten. Überdies erfordert eine kanonische Repräsentation risikobehafteter Gewinnaussichten eine Verknüpfung aller Ergebnisse konkurrierender Entscheidungen (z. B. Problem 4), die die Fähigkeiten intuitiver Berechnung selbst bei einfachen Problemen übersteigt. In anderen Kontexten wie Sicherheit, Gesundheit oder Lebensqualität lässt sich eine kanonische Repräsentation noch schwerer erreichen. Sollten wir Menschen raten, die Folgen einer gesundheitspolitischen Maßnahme (z. B. Probleme 1 und 2) nach der Gesamtsterblichkeit, der krankheitsbedingten Sterblichkeit oder der Anzahl der mit einer bestimmten Krankheit verbundenen Todesfälle zu beurteilen?

Eine zweite Vorgehensweise, die die Invarianz gewährleisten könnte, ist die Bewertung von Optionen anhand ihrer versicherungsmathematischen statt ihrer psychologischen Konsequenzen. Das versicherungsmathematische Kriterium besitzt einen gewissen Reiz, wenn es um Menschenleben geht, aber es ist eindeutig unzureichend für finanzielle Entscheidungen, wie zumindest seit Bernoulli allgemein anerkannt ist; und es lässt sich keinesfalls auf Ergebnisse anwenden, die keine objektive Metrik besitzen. Wir gelangen zu dem Schluss, dass die Framing-Invarianz nicht aufrechtzuerhalten ist und dass das subjektive Überzeugtsein von der Richtigkeit einer bestimmten Entscheidung nicht gewährleistet, dass bei einer anderen »Einrahmung« respektive Darstellung die gleiche Entscheidung getroffen würde. Es ist daher ratsam, die Robustheit von Präferenzen durch gezielte Versuche zu überprüfen, ein Entscheidungsproblem auf mehr als eine Weise zu formulieren.[11]

## Die Psychophysik von Wahrscheinlichkeiten

Unsere Diskussion ging bislang von einer bernoullischen Erwartungsregel aus, wonach man den Wert oder Nutzen einer unsicheren Aussicht dadurch erhält, dass man die Nutzwerte der möglichen Ergebnisse addiert, gewichtet nach ihrer jeweiligen Wahrscheinlichkeit. Zur Überprüfung dieser Annahme wollen wir ein weiteres Mal psychophysische Intuitionen zurate ziehen. Wir setzen den Wert des Status quo mit null an und stellen uns ein Geldgeschenk vor, zum Beispiel in Höhe von 300 Dollar, dem wir einen Wert von eins zuschreiben. Stellen wir uns jetzt vor, dass wir ein Lotterielos mit einem einzigen Gewinn in Höhe von 300 Dollar erhalten. Wie verändert sich der Wert des Lotteriescheins in Abhängigkeit von den Gewinnchancen? Sieht man von der subjektiven Wertschätzung des Glücksspiels ab, muss der Wert einer solchen Gewinnaussicht zwischen null (die Gewinnchancen sind null) und eins (wenn der Gewinn von 300 Dollar sicher ist) schwanken.

Die Intuition sagt uns, dass der Wert des Loses keine lineare Funktion der Gewinnwahrscheinlichkeit ist, wie aus der Erwartungsregel folgt. Insbesondere scheint eine Erhöhung von 0 auf 5 Prozent einen größeren Effekt zu haben als eine Zunahme von 30 auf 35 Prozent, die ihrerseits nicht so stark ins Gewicht fällt wie ein Anstieg von 95 auf 100 Prozent. Diese Erwägungen deuten auf einen Kategoriegrenzeffekt hin: Eine Veränderung von Unmöglichkeit zu Möglichkeit oder von Möglichkeit zu Gewissheit wirkt sich stärker aus als eine vergleichbare Veränderung in der Mitte der Skala. Diese Hypothese wird bei der in Abbildung 2 gezeigten Kurve berücksichtigt, die das mit einem Ereignis verbundene Gewicht als eine Funktion seiner angegebenen numerischen Wahrscheinlichkeit darstellt. Das auffälligste Merkmal von Abbildung 2 ist die Tatsache, dass Entscheidungsgewichte in Bezug auf die angegebenen Wahrscheinlichkeiten regressiv sind. Mit Ausnahme der Bereiche um die Endpunkte erhöht eine Zunahme von 0,05 in der Gewinnwahrscheinlichkeit den Wert der Aussicht um weniger als 5 Prozent des Werts der Gewinnprämie. Als Nächstes betrachten wir die Auswirkungen dieser psychophysischen Hypothesen auf Präferenzen bei risikobehafteten Optionen.

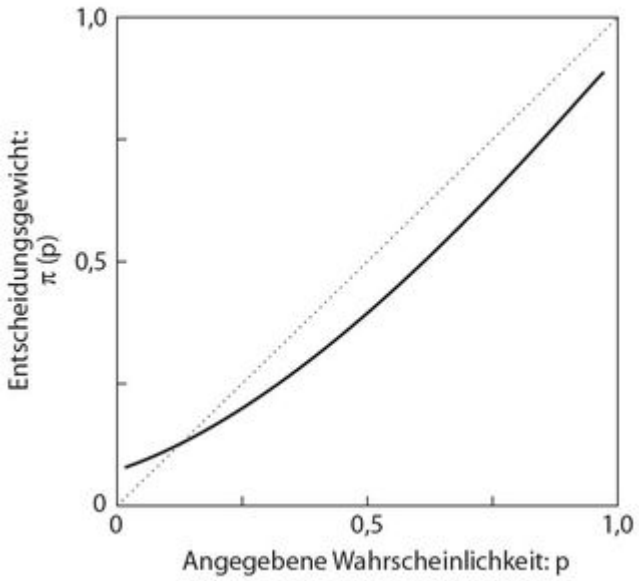


**Abbildung 2**
**Eine hypothetische Gewichtungsfunktion**

In Abbildung 2 sind die Entscheidungsgewichte über den größten Teil des Wertebereichs niedriger als die entsprechenden Wahrscheinlichkeiten. Die Untergewichtung mittlerer und hoher Wahrscheinlichkeiten im Vergleich zu sicheren Optionen trägt zur Risikoaversion bei Gewinnen bei, indem sie die Attraktivität positiver Lotterien verringert. Der gleiche Effekt trägt auch zur Risikoneigung bei Verlusten bei, indem er die Aversivität negativer Lotterien abschwächt. Niedrige Wahrscheinlichkeiten dagegen werden übergewichtet, und sehr niedrige Wahrscheinlichkeiten werden entweder stark übergewichtet oder gänzlich vernachlässigt, was die Entscheidungsgewichte in diesem Bereich höchst instabil macht. Die Übergewichtung niedriger Wahrscheinlichkeiten kehrt das oben beschriebene Muster um: Sie steigert den (subjektiven) Wert hochspekulativer Wetten und verstärkt die Aversivität einer geringen Wahrscheinlichkeit eines schweren Verlustes. Folglich sind Menschen oftmals risikofreudig, wenn sie es mit unwahrscheinlichen Gewinnen zu tun haben, und risikoscheu bei unwahrscheinlichen Verlusten. Somit tragen die Merkmale von Entscheidungsgewichten zur Attraktivität von Lotterielosen wie auch von Versicherungspolicen bei.

Die Nichtlinearität von Entscheidungsgewichten führt zwangsläufig zu Verletzungen der Invarianz, wie das folgende Paar von Problemen verdeutlicht:

Problem 5 (*N* = 85): Betrachten Sie das folgende zweiphasige Spiel. In der ersten Phase besteht eine 75-prozentige Chance, das Spiel ohne Gewinn zu beenden, und eine 25-prozentige Chance, in die zweite Phase zu gelangen. In der zweiten Phase hat man eine Wahl zwischen:

A Einem sicheren Gewinn von 30 Dollar. (74 Prozent)
B Einer 80-prozentigen Chance, 45 Dollar zu gewinnen. (26 Prozent)

Sie müssen sich entscheiden, bevor das Spiel beginnt, das heißt, bevor das Ergebnis der ersten Phase bekannt ist. Bitte geben Sie an, welche Option Sie vorziehen.

Problem 6 (*N* = 81): Welche der folgenden Optionen ziehen Sie vor?

C Eine 25-prozentige Chance, 30 Dollar zu gewinnen. (42 Prozent)
D Eine 20-prozentige Chance, 45 Dollar zu gewinnen. (58 Prozent)

Weil die Chance, die zweite Phase von Problem 5 zu erreichen, 1:4 beträgt, ist Aussicht A mit einer Wahrscheinlichkeit von 0,25 verbunden, 30 Dollar zu gewinnen, während Aussicht B eine Wahrscheinlichkeit von 0,25 × 0,80 = 0,20 bietet, 45 Dollar zu gewinnen. Daher sind die Probleme 5 und 6 in Bezug auf Wahrscheinlichkeiten und Ergebnisse identisch. Aber die Präferenzen sind in beiden Versionen nicht die gleichen: Eine klare Mehrheit bevorzugt bei Problem 5 die höhere Chance, den kleineren Betrag zu gewinnen, während sich die Mehrheit bei Problem 6 für das Gegenteil entscheidet. Diese Verletzung der Invarianz wurde sowohl bei realen wie bei hypothetischen Geldgewinnen (die gegenwärtigen Ergebnisse beziehen

sich auf reale Geldgewinne), bei Menschenleben als Ergebnissen und bei einer nicht sequenziellen Repräsentation des Zufallsprozesses beobachtet.

Wir führen die Verletzung der Invarianz auf die Wechselwirkung zweier Faktoren zurück: des Framings von Wahrscheinlichkeiten und der Nichtlinearität von Entscheidungsgewichten. Insbesondere nehmen wir an, dass Menschen bei Problem 5 die erste Phase, die unabhängig von der getroffenen Entscheidung das gleiche Ergebnis zeitigt, ignorieren und ihre Aufmerksamkeit auf das konzentrieren, was geschieht, wenn sie die zweite Phase des Spiels erreichen. In diesem Fall haben sie die Wahl zwischen einem sicheren Gewinn, wenn sie sich für Option A entscheiden, und einer 80-prozentigen Gewinnchance, wenn sie die Lotterie vorziehen. Tatsächlich sind die Entscheidungen der Versuchsteilnehmer in der sequenziellen Version praktisch identisch mit den Wahlen, die sie zwischen einem sicheren Gewinn von 30 Dollar und einer 85-prozentigen Chance, 45 Dollar zu gewinnen, trafen. Weil eine sichere Option im Vergleich mit Ereignissen mittlerer oder hoher Wahrscheinlichkeit übergewichtet wird (vgl. Abbildung 2), ist die Option, die möglicherweise zu einem Gewinn von 30 Dollar führt, in der sequenziellen Version attraktiver. Wir nennen dieses Phänomen den »Pseudo-Sicherheitseffekt«, weil ein Ereignis, das faktisch unsicher ist, so gewichtet wird, als wäre es sicher.

Ein eng verwandtes Phänomen lässt sich am unteren Ende des Wahrscheinlichkeitsbereichs nachweisen. Angenommen, Sie sind unentschieden, ob Sie eine Versicherung gegen Erdbeben abschließen sollen, weil die Prämie ziemlich hoch ist. Während Sie noch zögern, legt Ihnen Ihr freundlicher Versicherungsvertreter ein alternatives Angebot vor: »Für die Hälfte der regulären Prämie erhalten Sie einen vollumfänglichen Versicherungsschutz, wenn sich das Erdbeben an einem ungeraden Tag des Monats ereignet. Dies ist ein gutes Geschäft, denn für die Hälfte des Preises erhalten Sie für mehr als die Hälfte der Tage Versicherungsschutz.« Weshalb finden die meisten Menschen solche probabilistischen Versicherungspolicen ausgesprochen unattraktiv? Abbildung 2 legt eine Antwort nahe. Irgendwo im Bereich niedriger Wahrscheinlichkeiten beginnend, wirkt sich eine Verringerung der Wahrscheinlichkeit von p zu p/2 in viel geringerem Maße auf das Entscheidungsgewicht aus als eine Reduktion von p/2 zu 0. Die Halbierung des Risikos ist daher nicht die Hälfte der Prämie wert.

Die Abneigung gegen probabilistische Versicherungen ist aus drei Gründen bedeutsam. Erstens untergräbt sie die klassische Erklärung von Versicherungen auf der Basis einer konkaven Nutzenfunktion. Gemäß der Erwartungsnutzentheorie sollte eine probabilistische Versicherung definitiv einer normalen Versicherung vorgezogen werden, wenn Letztere gerade akzeptabel ist.[12] Zweitens repräsentieren probabilistische Versicherungen viele Formen von Schutzmaßnahmen, wie etwa eine ärztliche Vorsorgeuntersuchung, den Kauf neuer Reifen oder den Einbau einer Alarmanlage. Solche Maßnahmen verringern typischerweise die Wahrscheinlichkeit eines Schadensereignisses, ohne das Risiko vollkommen auszuschalten. Drittens kann die Akzeptanz von Versicherungen durch das Framing ungewisser zukünftiger Schadensereignisse gezielt beeinflusst werden. Eine Versicherungspolice, die Brand-, nicht aber Hochwasserschäden abdeckt, könnte zum Beispiel entweder als ein vollumfänglicher Versicherungsschutz gegen ein spezifisches Risiko (zum Beispiel Feuer) oder als eine Verringerung der Gesamtwahrscheinlichkeit von Sachschäden beurteilt werden. Abbildung 2 deutet darauf hin, dass Menschen eine Verringerung der Wahrscheinlichkeit eines Schadensereignisses im Vergleich zur vollständigen Beseitigung dieses Risikos deutlich unterbewerten. Folglich sollte eine Versicherung attraktiver erscheinen, wenn sie so dargestellt wird, als würde sie das Risiko vollkommen beseitigen, als wenn sie laut ihrer Beschreibung das Risiko nur verringern würde. Tatsächlich haben Slovic, Fischhoff und Lichtenstein[13] gezeigt, dass ein hypothetischer Impfstoff, der die Wahrscheinlichkeit, sich eine Infektionskrankheit zuzuziehen, von 20 auf 10 Prozent verringert, weniger attraktiv ist, wenn er als in der Hälfte der Fälle wirksam beschrieben wird, als wenn er so dargestellt wird, dass er gegen einen von zwei exklusiven und gleich wahrscheinlichen Virusstämmen mit identischen Symptomen hundertprozentig wirksam ist.

## Formulierungseffekte

Bis jetzt haben wir Framing als ein Werkzeug diskutiert, mit dem sich Verstöße gegen das Invarianzaxiom veranschaulichen lassen. Jetzt wenden wir unsere Aufmerksamkeit den Prozessen zu, die das Framing von Ergebnissen und Ereignissen kontrollieren. Das gesundheitspolitische Problem veranschaulicht einen Formulierungseffekt, bei dem eine Veränderung der Wortwahl von »gerettete Menschenleben« in »Verluste an Menschenleben« eine deutliche Verschiebung der Präferenz von Risikoscheu zu Risikoneigung auslöste. Offensichtlich nahmen die Versuchsteilnehmer die Beschreibungen der Ergebnisse so hin, wie sie in der Frage vorgegeben wurden, und beurteilten die Ergebnisse entsprechend als Gewinne oder Verluste. Ein weiterer Formulierungseffekt wurde von McNeil, Pauker, Sox und Tversky[14] berichtet. Sie fanden heraus, dass Präferenzen von Ärzten und Patienten für hypothetische Lungenkrebstherapien deutlich schwankten, je nachdem, ob ihre wahrscheinlichen Ergebnisse in Form von Sterblichkeit oder Überlebensrate beschrieben wurden. Anders als die Strahlentherapie sind Operationen mit einem Sterberisiko während der Behandlung verbunden. Infolgedessen war die chirurgische Option vergleichsweise weniger attraktiv, wenn die Statistiken der Behandlungsergebnisse in Kategorien der Sterblichkeit statt in Kategorien von Überlebensraten beschrieben wurden.

Ein Arzt und vielleicht auch ein Präsidentenberater könnte die Entscheidung, die der Patient beziehungsweise der Präsident trifft, allein durch die »Einrahmung« – die Darstellungsweise – von Ergebnissen und Risiken beeinflussen, ohne Informationen zu verzerren oder zu unterdrücken. Formulierungseffekte können zufällig auftreten, ohne dass sich irgendjemand des Einflusses der Formulierung auf die letztendlich getroffene Entscheidung bewusst ist. Sie können auch bewusst ausgenutzt werden, um die relative Attraktivität von Optionen zu manipulieren. So wies beispielsweise Thaler[15] darauf hin, dass Lobbyisten der Kreditkartenbranche darauf bestanden, jeden Preisunterschied zwischen Bar- und Kreditkäufen als Barrabatt statt als Kreditkartenaufschlag zu bezeichnen. Die beiden Bezeichnungen »rahmen« den Preisunterschied als einen Gewinn oder einen Verlust, indem sie implizit entweder den höheren

oder den niedrigeren Preis als normal kennzeichnen. Weil Verluste schwerer wiegen als Gewinne, akzeptieren Konsumenten mit geringerer Wahrscheinlichkeit einen Zuschlag, als dass sie sich einen Rabatt entgehen lassen. Wie zu erwarten, sind Versuche, durch Framing zu manipulieren, sowohl am Markt als auch in der politischen Arena weit verbreitet.

Die Bewertung von Ergebnissen ist wegen der Nichtlinearität der Wertfunktion und der Tendenz von Menschen, Optionen in Bezug auf den Referenzpunkt zu beurteilen, der durch die Darlegung des Problems nahegelegt oder angedeutet wird, anfällig für Formulierungseffekte. Es sei darauf hingewiesen, dass Menschen in anderen Kontexten automatisch äquivalente Botschaften in die gleiche Repräsentation transformieren. Studien über Sprachverstehen deuten darauf hin, dass Menschen einen Großteil dessen, was sie hören, in eine abstrakte Repräsentation umkodieren, die nicht länger unterscheidet, ob der Gedanke in einer aktiven oder passiven Form ausgedrückt wurde, und auch keinen Unterschied mehr macht zwischen dem, was tatsächlich gesagt wurde, und dem, was angedeutet, stillschweigend vorausgesetzt oder impliziert wurde.[16] Leider reicht die mentale Maschinerie, die diese Operationen still und mühelos ausführt, nicht aus, um die Aufgabe zu bewältigen, die beiden Versionen des gesundheitspolitischen Problems beziehungsweise der Sterblichkeits-Überlebensrate-Statistik in eine gemeinsame abstrakte Form umzukodieren.

## Transaktionen und Tauschgeschäfte

Unsere Analyse von Framing und subjektivem Wert lässt sich auf Wahlen zwischen multiattributiven Optionen ausdehnen, wie etwa der Annehmbarkeit einer Transaktion oder eines Tausches. Zur Bewertung einer multiattributiven Funktion schlagen wir vor, dass eine Person ein mentales Konto einrichtet, das die Vor- und Nachteile, die mit dieser Option verbunden sind, in Bezug auf einen multiattributiven Referenzzustand genau verbucht. Der Gesamtwert einer Option ist gegeben durch den Saldo ihrer Vor- und Nachteile in Bezug auf den Referenzzustand. Folglich ist eine Option dann annehmbar, wenn der Wert ihrer Vorteile den Wert ihrer Nachteile überwiegt. Diese Analyse geht von der Annahme aus, dass sich Vor- und Nachteile psychologisch – nicht jedoch physisch – voneinander trennen lassen. Das Modell beschränkt nicht die Art und Weise, in der getrennte Attribute kombiniert werden können, um globale Maße der Vor- und Nachteile zu bilden, aber es erlegt diesen Maßen Konkavitäts- und Verlustaversionsannahmen auf.

Unsere Analyse der mentalen Buchführung verdankt sehr viel den anregenden Arbeiten von Richard Thaler,[17] der die Relevanz dieses Prozesses für das Verbraucherverhalten nachwies. Das folgende Problem, das auf Beispielen von Savage[18] und Thaler[19] basiert, führt einige der Regeln ein, denen die Konstruktion mentaler Konten gehorcht, und veranschaulicht die Erweiterung der Konkavität des Wertes auf die Akzeptabilität von Transaktionen.

> Problem 7: Stellen Sie sich vor, dass Sie eine Jacke für 125 Dollar und einen Taschenrechner für 15 Dollar kaufen wollen. Der Verkäufer des Taschenrechners informiert Sie, dass der Rechner, den Sie kaufen wollen, in einer andere Filiale der Kette, die etwa zwanzig Fahrminuten entfernt liegt, für 10 Dollar im Sonderangebot ist. Würden Sie zu der anderen Filiale fahren?

Das Problem betrifft die Annehmbarkeit einer Option, die einen Nachteil (Unannehmlichkeit) mit einem finanziellen Vorteil verbindet, der als minimales, thematisches (»topisches«) oder umfassendes Konto gerahmt werden kann. Das minimale Konto erfasst lediglich die Unterschiede zwischen den beiden Optionen und lässt ihre gemeinsamen Merkmale außer Betracht. Beim minimalen Konto wird der Vorteil, der mit der Fahrt zu der anderen Filiale verbunden ist, als ein Gewinn von 5 Dollar verbucht. Ein thematisches Konto bezieht die Folgen möglicher Wahlen auf ein Referenzniveau, das durch den Kontext bestimmt wird, innerhalb dessen die Entscheidung stattfindet. Beim beschriebenen Problem ist das relevante Thema der Kauf des Taschenrechners, und der Nutzen der Fahrt wird daher als ein Preisnachlass von 15 auf 10 Dollar formuliert. Weil die potenzielle Ersparnis nur mit dem Taschenrechner verknüpft ist, wird der Preis der Jacke nicht auf dem thematischen Konto verbucht. Der Preis der Jacke könnte allerdings, wie andere Ausgaben, auf einem umfassenderen Konto vermerkt werden, auf dem die Ersparnis zum Beispiel im Verhältnis zu den monatlichen Ausgaben bewertet wird.

Die Formulierung des voranstehenden Problems erscheint neutral im Hinblick auf die Verwendung eines minimalen, thematischen oder umfassenden Kontos. Wir behaupten jedoch, dass Menschen spontan Entscheidungen in Form von thematischen Konten erfassen, die, im Kontext der Entscheidungsfindung, eine ähnliche Rolle spielen wie die »gute Gestalt« in der Wahrnehmung und Basiskategorien in der Kognition. Eine thematische Organisation hat in Verbindung mit der Konkavität der Wertfunktion zur Folge, dass die Bereitschaft, zu einer anderen Filiale zu fahren, um einen Taschenrechner 5 Dollar billiger zu kaufen, in umgekehrtem Verhältnis zum Preis des Taschenrechners stehen und unabhängig vom Preis der Jacke sein sollte. Um diese Vorhersage zu überprüfen, entwarfen wir eine andere Version des Problems, in dem die Preise der beiden Artikel vertauscht wurden. Der Preis des Taschenrechners wurde im ersten Geschäft mit 125 Dollar und in der zweiten Filiale mit 120 Dollar angegeben, und der Preis der Jacke wurde auf 15 Dollar festgesetzt. Wie vorhergesagt, wichen die Prozentsätze der Befragten, die sagten, sie würden die Fahrt machen, bei den beiden Problemen erheblich voneinander ab. Die Ergebnisse zeigten, dass 68 Prozent der Befragten *(N = 88)* gewillt waren, zu der anderen Filiale zu fahren, um 5 Dollar für einen Taschenrechner mit einem Preis von 15 Dollar zu sparen, aber nur 29

Prozent der 93 Befragten wollten die Fahrt machen, um bei einem Rechner, der 125 Dollar kosten sollte, 5 Dollar zu sparen. Dieses Ergebnis bekräftigt die Hypothese einer thematischen Organisation mentaler Konten, da die beiden Versionen hinsichtlich ihrer Erfassung auf einem minimalen und einem umfassenden Konto identisch sind.

Die Bedeutung thematischer Konten für das Verbraucherverhalten wird durch die Beobachtung bestätigt, dass die Standardabweichung der Preise, die verschiedene Geschäfte in einer Innenstadt für das gleiche Produkt angeben, annähernd proportional zum Durchschnittspreis dieses Produktes ist.[20] Da die Streuung der Preise sicherlich durch die Bemühungen der Käufer kontrolliert wird, das günstigste Schnäppchen zu finden, deuten diese Ergebnisse darauf hin, dass Verbraucher sich kaum mehr anstrengen dürften, um bei einem Kauf zu 150 Dollar 15 Dollar zu sparen, als bei einem Kauf zu 50 Dollar 5 Dollar zu sparen.

Die thematische Organisation mentaler Konten veranlasst Menschen dazu, Gewinne und Verluste eher relativ als absolut zu beurteilen, sodass es große Variationen in der Rate gibt, mit der Geld gegen andere Dinge getauscht wird, wie etwa die Anzahl der Anrufe, die getätigt werden, um einen guten Kauf zu machen, oder die Bereitschaft, dafür eine weite Strecke zu fahren. Den meisten Verbrauchern fällt es leichter, in Verbindung mit einem Auto- oder Hauskauf – als losgelöst davon – eine Auto-Stereoanlage oder einen persischen Teppich zu kaufen. Diese Beobachtungen widersprechen natürlich der herrschenden rationalen Theorie des Verbraucherverhaltens, die von der Annahme der Invarianz ausgeht und die Effekte der mentalen Buchführung nicht anerkennt.

Die folgenden Probleme verdeutlichen ein weiteres Beispiel der mentalen Buchführung, in dem die Verbuchung von Kosten auf einem Konto einer thematischen Organisation folgt:

> Problem 8 ($N$ = 200): Stellen Sie sich vor, Sie haben beschlossen, sich ein Theaterstück anzusehen, und Sie haben den Eintrittspreis von 10 Dollar pro Karte bezahlt. Beim Betreten des Theaters bemerken Sie, dass Sie die Karte verloren haben. Der Sitzplatz war nicht markiert, und der Eintrittspreis wird nicht erstattet.

Würden Sie für eine zweite Karte 10 Dollar bezahlen?
Ja (46 Prozent) – Nein (54 Prozent)

Problem 9 (*N* = 183): Stellen Sie sich vor, Sie haben beschlossen, sich ein Theaterstück anzusehen, bei dem die Eintrittskarte 10 Dollar kostet. Beim Betreten des Theaters stellen Sie fest, dass Sie einen Zehn-Dollar-Schein verloren haben.

Würden Sie nach wie vor 10 Dollar für eine Eintrittskarte bezahlen?
Ja (88 Prozent) – Nein (12 Prozent)

Der Unterschied zwischen den Antworten auf die beiden Probleme ist verblüffend. Weshalb sind so viele Menschen nicht bereit, nach dem Verlust einer Eintrittskarte 10 Dollar auszugeben, während sie diese Summe nach dem Verlust eines gleich hohen Bargeldbetrags ohne Weiteres ausgeben würden? Wir führen den Unterschied auf die thematische Organisation mentaler Konten zurück. Ein Theaterbesuch wird normalerweise als eine Transaktion betrachtet, bei der die Kosten der Karte gegen die Erfahrung, der Aufführung eines Stückes beizuwohnen, getauscht werden. Der Kauf einer zweiten Karte erhöht die Kosten für das Sehen des Stücks auf ein Niveau, das viele Befragte offensichtlich unannehmbar finden. Dagegen wird der Verlust von Bargeld nicht auf dem Konto des Stückes verbucht, und er wirkt sich nur insoweit auf den Kauf einer Eintrittskarte aus, als er dem Individuum das Gefühl vermittelt, etwas weniger wohlhabend zu sein.

Ein interessanter Effekt wurde beobachtet, als die beiden Versionen des Problems denselben Probanden dargeboten wurden. Die Bereitschaft, eine verlorene Eintrittskarte zu ersetzen, nahm deutlich zu, wenn das Problem der Version des verlorenen Geldes folgte. Dagegen wurde die Bereitschaft, nach dem Verlust von Bargeld eine Eintrittskarte zu kaufen, nicht von der vorhergehenden Präsentation des anderen Problems beeinflusst. Die Nebeneinanderstellung der beiden Probleme befähigte die Probanden

offensichtlich, zu erkennen, dass es sinnvoll ist, die verlorene Eintrittskarte als verlorenes Geld aufzufassen, nicht aber umgekehrt.

Der normative Status der Effekte mentaler Buchführung ist fragwürdig. Anders als bei früheren Beispielen, wie dem gesundheitspolitischen Problem, bei dem die beiden Versionen sich nur formal unterschieden, kann man behaupten, dass sich die alternativen Versionen der Taschenrechner- und Eintrittskarten-Probleme auch in der Sache unterscheiden. So mag es insbesondere mit einem stärkeren Lustgefühl verbunden sein, bei einem 15-Dollar-Kauf 5 Dollar zu sparen, als bei einem teureren Kauf, und es mag ärgerlicher sein, zweimal für die gleiche Karte zu bezahlen, als 10 Dollar in bar zu verlieren. Auch Reue, Frustration und Selbstzufriedenheit können durch Framing beeinflusst werden.[21] Wenn solche sekundären Konsequenzen als legitim betrachtet werden, dann verletzen die beobachteten Präferenzen nicht das Kriterium der Invarianz und können nicht ohne Weiteres als inkonsistent oder fehlerhaft verworfen werden. Andererseits mögen sich die sekundären Konsequenzen bei genauerem Nachdenken verändern. Die Befriedigung darüber, bei einem 15 Dollar teuren Artikel 5 Dollar zu sparen, kann gemindert werden, wenn der Verbraucher entdeckt, dass er sich nicht die gleiche Mühe gegeben hätte, um bei einem 200-Dollar-Kauf 10 Dollar zu sparen. Wir empfehlen nicht, dass jedwede zwei Entscheidungsprobleme, die die gleichen primären Konsequenzen haben, in der gleichen Weise gelöst werden. Wir behaupten jedoch, dass die systematische Prüfung alternativer Framings ein nützliches gedankliches Werkzeug darstellt, das Entscheidern helfen kann, die Werte zu beurteilen, die den primären und sekundären Konsequenzen ihrer Wahlen beigefügt werden sollten.

## Verluste und Kosten

Viele Entscheidungsprobleme nehmen die Form einer Wahl zwischen der Beibehaltung des Status quo und des Akzeptierens einer Alternative an, die in mancher Hinsicht vorteilhaft und in anderer Hinsicht nachteilig ist. Die Analyse des Wertes, die früher auf eindimensionale riskante Gewinnaussichten angewandt wurde, lässt sich auf diesen Fall erweitern, indem man annimmt, dass der Status quo das Referenzniveau für alle Attribute definiert. Die Vorteile alternativer Optionen werden dann als Gewinne bewertet und ihre Nachteile als Verluste. Weil Verluste schwerer wiegen als Gewinne, wird der Entscheider zugunsten der Beibehaltung des Status quo voreingenommen sein.

Thaler[22] prägte den Begriff »Endowment-Effekt« (»Besitztumseffekt«), um die Tatsache zu beschreiben, dass sich Menschen nur widerstrebend von ihren Vermögenswerten trennen. Wenn es schmerzlicher ist, einen Vermögenswert aufzugeben, als es erfreulich ist, einen solchen zu erlangen, werden die Kaufpreise deutlich niedriger sein als die Verkaufspreise. Das heißt, der höchste Preis, den ein Individuum zu zahlen bereit ist, um einen Vermögenswert zu erwerben, ist geringer als die Mindestentschädigung, die dasselbe Individuum dazu veranlassen würde, diesen Vermögenswert aufzugeben, nachdem es ihn erworben hat. Thaler diskutierte einige Beispiele des Endowment-Effekts im Verhalten von Verbrauchern und Unternehmern. Mehrere Studien haben über deutliche Diskrepanzen zwischen Kauf- und Verkaufspreisen bei hypothetischen und realen Transaktionen berichtet.[23] Diese Ergebnisse wurden als Herausforderungen für die herrschende volkswirtschaftliche Theorie dargestellt, der zufolge sich Kauf- und Verkaufspreise decken, einmal abgesehen von den Transaktionskosten und Vermögenseffekten. Wir beobachteten auch eine »Abneigung gegen Tausch«, und zwar in einer Studie über Wahlen zwischen hypothetischen Arbeitsplätzen, die sich im wöchentlichen Gehalt (S) und in der Temperatur (T) am Arbeitsplatz unterschieden. Wir baten unsere Probanden, sich vorzustellen, sie würden eine bestimmte Position ($S_1$, $T_1$) bekleiden und erhielten die Option, auf eine andere Stelle ($S_2$, $T_2$) zu wechseln, die in einer Hinsicht besser und in einer anderen Hinsicht schlechter war. Wir stellten fest, dass die meisten Versuchspersonen, die

($S_1$, $T_1$) zugewiesen wurden, nicht auf ($S_2$, $T_2$) wechseln wollten, und dass die meisten Probanden, denen die zweite Stelle zugewiesen wurde, nicht auf die erste wechseln wollten. Offensichtlich wiegt der gleiche Unterschied im Gehalt oder in den Arbeitsbedingungen als Nachteil schwerer denn als Vorteil.

Im Allgemeinen begünstigt die Verlustaversion Stabilität statt Veränderungen. Stellen Sie sich zwei hedonisch identische Zwillinge vor, die zwei alternative Umgebungen gleich attraktiv finden. Stellen Sie sich des Weiteren vor, dass die Zwillinge, bedingt durch die Umstände, getrennt werden und in den beiden Umgebungen untergebracht werden. Sobald sie ihre neuen Zustände als Referenzpunkte übernehmen und die Vor- und Nachteile ihrer jeweiligen Umgebungen entsprechend bewerten, werden die Zwillinge gegenüber den beiden Zuständen nicht länger indifferent sein, und beide werden es vorziehen, dort zu bleiben, wo sie sich aufhalten. Die Instabilität von Präferenzen erzeugt folglich eine Präferenz für Stabilität. Die Kombination von Anpassung und Verlustaversion begünstigt nicht nur Stabilität statt Veränderung, sie bietet auch einen gewissen Schutz gegen Reue und Neid, indem sie die Attraktivität entgangener Alternativen und des Besitzes anderer Personen verringert.

Die Verlustaversion und der daraus folgende Endowment-Effekt dürften bei routinemäßigen ökonomischen Tauschgeschäften keine bedeutende Rolle spielen. Der Besitzer eines Ladens zum Beispiel erlebt Geld, das er an Lieferanten zahlt, nicht als Verlust und Geld, das er von Kunden erhält, nicht als Gewinn. Vielmehr addiert der Kaufmann Kosten und Erlöse über einen bestimmten Zeitraum und beurteilt nur den Saldo. Übereinstimmende Soll- und Habenposten werden vor der Beurteilung faktisch gestrichen. Zahlungen von Verbrauchern werden ebenfalls nicht als Verluste beurteilt, sondern als alternative Käufe. In Übereinstimmung mit der gängigen ökonomischen Analyse wird Geld als ein Stellvertreter für Güter und Dienstleistungen betrachtet, die mit ihm gekauft werden könnten. Diese Beurteilungsweise wird deutlich, wenn eine Person an eine bestimmte Alternative denkt, wie etwa: »Ich kann entweder eine neue Kamera oder ein neues Zelt kaufen.« Nach dieser Analyse wird eine Person eine Kamera kaufen, wenn deren subjektiver Wert den Wert übersteigt, den das Behalten des Geldes, das sie kostet, besitzt.

Es gibt Fälle, in denen ein Nachteil entweder als ein Kostenfaktor oder als ein Verlust ausgedrückt werden kann. Insbesondere kann das

Abschließen einer Versicherung ebenfalls als eine Wahl zwischen einem sicheren Verlust und dem Risiko eines größeren Verlusts beschrieben werden. In solchen Fällen kann die Kosten-Verlust-Diskrepanz zu Verletzungen der Invarianz führen. Betrachten wir zum Beispiel die Wahl zwischen einem sicheren Verlust von 50 Dollar und einer 25-prozentigen Chance, 200 Dollar zu verlieren. Slovic, Fischhoff und Lichtenstein [24] berichteten, dass 80 Prozent ihrer Probanden eine risikogeneigte Präferenz für die Lotterie gegenüber dem sicheren Verlust bekundeten. Doch nur 35 Prozent der Probanden wollten keine 50 Dollar für eine Versicherung gegen ein 25-prozentiges Risiko, 200 Dollar zu verlieren, bezahlen. Ähnliche Ergebnisse berichteten Schoemaker und Kunreuther[25] sowie Hershey und Schoemaker [26]. Wir meinen, dass der gleiche Geldbetrag, der im ersten Problem als nicht ausgeglichener Verlust beschrieben wurde, im zweiten Problem als Kosten der Absicherung gerahmt wurde. Die modale Präferenz wurde bei den beiden Problemen umgekehrt, weil Verluste eine stärkere Aversion hervorrufen als Kosten.

Im positiven Bereich haben wir einen ähnlichen Effekt beobachtet, wie das folgende Paar von Problemen verdeutlicht:

> Problem 10: Würden Sie eine Lotterie akzeptieren, die eine 10-prozentige Chance bietet, 95 Dollar zu gewinnen, und eine 90-prozentige Chance, 5 Dollar zu verlieren?

> Problem 11: Würden Sie 5 Dollar für die Teilnahme an einer Lotterie bezahlen, die eine 10-prozentige Chance bietet, 100 Dollar zu gewinnen, und eine 90-prozentige Chance, nichts zu gewinnen?

Insgesamt beantworteten 132 Studenten die beiden Fragen, die durch ein kurzes Zwischenproblem getrennt wurden. Für die Hälfte der Befragten wurde die Reihenfolge der Fragen umgekehrt. Obgleich sich leicht bestätigen lässt, dass die beiden Probleme objektiv identische Optionen

bieten, äußerten 55 Befragte in den beiden Versionen unterschiedliche Präferenzen. 42 von ihnen lehnten die Lotterie in Problem 10 ab, akzeptierten jedoch die äquivalente Lotterie in Problem 11. Die Effektivität dieser scheinbar belanglosen Manipulation verdeutlicht sowohl die Kosten-Verlust-Diskrepanz als auch die Macht des Framings. Das Glücksspiel wird akzeptabler, wenn die 5 Dollar als Zahlung aufgefasst werden, als wenn der gleiche Betrag als Verlust betrachtet wird.

Aus der vorangehenden Analyse folgt, dass sich der subjektive Zustand eines Individuums dadurch verbessern lässt, dass man negative Ergebnisse als Kosten statt als Verluste darstellt. Die Möglichkeit solcher psychologischen Manipulationen mag eine paradoxe Form des Verhaltens erklären, die als »Totalverlust-Effekt« bezeichnet werden könnte. Thaler[27] diskutierte das Beispiel eines Mannes, der, bald nachdem er den Mitgliedsbeitrag in einem Tennisklub bezahlt hatte, einen Tennisarm entwickelte und unter starken Schmerzen weiterspielte, damit seine Ausgabe nicht umsonst war. In der Annahme, dass das Individuum nicht spielen würde, wenn es den Mitgliedsbeitrag nicht bezahlt hätte, erhebt sich folgende Frage: Wie kann das Spielen unter starken Schmerzen dem Wohlergehen des Individuums förderlich sein? Das Spielen unter Schmerzen, so meinen wir, erhält die Bewertung des Mitgliedsbeitrags als einen Kostenfaktor aufrecht. Wenn die Person mit dem Spielen aufhören würde, müsste sie den Beitrag als Totalverlust abschreiben, was eine stärkere Aversion hervorrufen mag, als unter Schmerzen zu spielen.

## Abschließende Bemerkungen

Die Begriffe »Nutzen« und »Wert« werden üblicherweise in zwei verschiedenen Bedeutungen gebraucht: (a) Erfahrungswert, das Ausmaß von Lust oder Unlust, Zufriedenheit oder Leid im tatsächlichen Erleben eines Ergebnisses; und (b) Entscheidungswert, der Beitrag eines antizipierten Ergebnisses zur Gesamtattraktivität oder -aversivität einer Option in einer Wahl. Die Unterscheidung wird in der Entscheidungstheorie selten explizit dargelegt, weil stillschweigend davon ausgegangen wird, dass Entscheidungswerte und Erfahrungswerte zusammenfallen. Diese Annahme ist Teil der Konzeption eines idealisierten Entscheiders, der zukünftige Erfahrungen vollkommen treffsicher vorhersagen und Optionen entsprechend bewerten kann. Für gewöhnliche Entscheider ist die Entsprechung von Entscheidungs- und Erfahrungswerten jedoch alles andere als perfekt.[28] Einige Faktoren, die die Erfahrung beeinflussen, lassen sich nicht leicht vorhersehen, und einige Faktoren, die Entscheidungen beeinflussen, haben keinen vergleichbaren Einfluss auf das Erleben der Ergebnisse.

Während sich eine große Zahl von Forschungsarbeiten mit der Entscheidungsfindung befassen, gibt es vergleichsweise wenige systematische Untersuchungen zur Psychophysik, in denen hedonisches Erleben mit objektiven Zuständen in Beziehung gesetzt wird. Das grundlegendste Problem der hedonischen Psychophysik ist die Bestimmung des Anpassungs- oder Anspruchsniveaus, das positive von negativen Ergebnissen trennt. Der hedonische Referenzpunkt ist weitgehend durch den objektiven Status quo definiert, aber er wird auch von Erwartungen und sozialen Vergleichen beeinflusst. So kann zum Beispiel eine objektive Verbesserung als ein Verlust erlebt werden, wenn ein Mitarbeiter eine geringere Gehaltserhöhung als alle anderen Kollegen im Büro erhält. Das Erleben von Lust oder Unlust, das mit einer Zustandsänderung verbunden ist, wird ebenfalls maßgeblich von der Dynamik hedonischer Anpassung beeinflusst. Das Konzept der hedonischen Tretmühle von Brickman und Campbell [29] legt die radikale Hypothese nahe, dass rasche Anpassung dafür sorgt, dass die Effekte jedweder objektiven Verbesserung von kurzer Dauer sein werden. Die Komplexität und Subtilität hedonischen Erlebens machen

es für den Entscheider schwer, die tatsächliche Erfahrung vorwegzunehmen, die Ergebnisse erzeugen werden. Viele Personen, die ein Menü bestellten, als sie einen Bärenhunger hatten, räumten nach dem fünften Gang ein, dass sie einen großen Fehler gemacht haben. Die weitverbreitete Diskrepanz zwischen Entscheidungs- und Erfahrungswerten führt in viele Entscheidungsprobleme ein zusätzliches Element der Unsicherheit ein.

Die Häufigkeit von Framing-Effekten und Verstößen gegen die Invarianz macht die Beziehung zwischen Entscheidungs- und Erfahrungswerten noch komplizierter. Die Einrahmung von Ergebnissen induziert oftmals Entscheidungswerte, die im tatsächlichen Erleben keine Entsprechung haben. So wird sich zum Beispiel die Darstellung der Ergebnisse von Therapien für Lungenkrebs nach Sterblichkeit oder Überlebensrate kaum auf das subjektive Erleben auswirken, während es einen deutlichen Einfluss auf die Wahl haben kann. In anderen Fällen jedoch beeinflusst die Einrahmung von Entscheidungen nicht nur die Entscheidung, sondern auch die Erfahrung. So kann zum Beispiel die Beschreibung einer Ausgabe als ein nicht ausgeglichener Verlust oder als Preis einer Versicherung wahrscheinlich das Erleben dieses Ergebnisses beeinflussen. In solchen Fällen nimmt die Bewertung von Ergebnissen im Kontext von Entscheidungen die Erfahrung nicht nur vorweg, sondern formt sie auch.

# *Dank*

Ich habe das Glück, viele Freunde zu haben und mich nicht dafür zu schämen, um Hilfe zu bitten. Ich habe mich an jeden meiner Freunde gewandt, an manche viele Male, und sie um Informationen oder redaktionelle Ratschläge gebeten. Ich bitte um Entschuldigung dafür, dass ich sie nicht alle aufführen kann. Einige Personen spielten eine wichtige Rolle in der Entstehungsgeschichte dieses Buches. Als Erstes möchte ich Jason Zweig danken, der mich zu dem Projekt drängte und sich geduldig darum bemühte, mit mir zu arbeiten, bis wir beide einsahen, dass dies ein Ding der Unmöglichkeit ist. Die ganze Zeit über hat er mir großzügig mit redaktionellen Ratschlägen und seiner beneidenswerten Bildung zur Seite gestanden, und zahlreiche Formulierungen, die er mir vorschlug, finden sich in dem Buch wieder. Roger Lewin arbeitete Vorlesungsprotokolle zu Rohfassungen von Kapiteln aus. Mary Himmelstein verdanke ich viele nützliche Anregungen. John Brockman wurde von einem Agenten zu einem getreuen Freund. Ran Hassin stand mir immer dann, wenn ich es am dringendsten brauchte, mit Rat und Ermunterung zur Seite. Am Ende eines langen Weges profitierte ich von der unschätzbaren Hilfe von Eric Chinski, meinem Lektor bei Farrar, Straus & Giroux. Er kannte das Buch besser als ich, und wir bildeten ein perfektes Gespann – Eric hat sich unglaublich für dieses Projekt ins Zeug gelegt. Meine Tochter, Lenore Shoham, unterstützte mich nach Kräften in den hektischen Schlussmonaten; ich profitierte von ihren klugen Ratschlägen und ihrem scharfen kritischen Auge, außerdem hat sie zahlreiche der Beispielsätze in den »Zum Thema«-Abschnitten beigesteuert. Meine Frau, Anne Treisman, hat eine Menge durchgemacht und getan – ohne ihre stetige Unterstützung, ihre Klugheit und endlose Geduld hätte ich schon vor langer Zeit aufgegeben.

# *Anmerkungen*

## Einleitung

1
Wir hatten ein Buch gelesen, das Psychologen dafür kritisierte, dass sie kleine Stichproben benutzten, aber ihre Optionen nicht darlegten: Jacob Cohen, *Statistical Power Analysis for the Behavioral Sciences*, Hillsdale 1969.

2
Ich habe den ursprünglichen Wortlaut leicht verändert, der sich auf die Buchstaben an der ersten und dritten Stelle innerhalb von Wörtern bezog.

3
Ein bekannter deutscher Psychologe ist unser beharrlichster Kritiker. Gerd Gigerenzer, »How to Make Cognitive Illusions Disappear«, *European Review of Social Psychology* 2 (1991): S. 83–115; Gerd Gigerenzer, »Personal Reflections on Theory and Psychology«, *Theory & Psychology* 20 (2010): S. 733–743. Daniel Kahneman und Amos Tversky, »On the Reality of Cognitive Illusions«, *Psychological Review* 103 (1996): S. 582–591.

4
Einige Beispiele von vielen sind Valerie F. Reyna und Farrell J. Lloyd, »Physician Decision-Making and Cardiac Risk: Effects of Knowledge, Risk Perception, Risk Tolerance and Fuzzy Processing«, *Journal of Experimental Psychology: Applied* 12 (2006): S. 179–195. Nicholas Epley and Thomas Gilovich, »The Anchoring-and-Adjustment Heuristic«, *Psychological Science* 17 (2006): S. 311–318. Norbert Schwarz et al., »Ease of Retrieval of Information: Another Look at the Availability Heuristic«, *Journal of Personality and Social Psychology* 61 (1991): S. 195–202. Elke U. Weber et al., »Asymmetric Discounting in Intertemporal Choice«, *Psychological Science* 18 (2007): S. 516–523. George F. Loewenstein et al., »Risk as Feelings«, *Psychological Bulletin* 127 (2001): S. 267–286.

5
Der in den Wirtschaftswissenschaften vergebene Nobelpreis heißt offiziell »Preis für Wirtschaftswissenschaften der schwedischen Reichsbank im

Gedenken an Alfred Nobel«. Er wurde erstmals 1969 vergeben. Einigen Naturwissenschaftlern missfiel es, dass zusätzlich ein Nobelpreis in den Sozialwissenschaften vergeben wurde; die besondere Bezeichnung des Preises für Wirtschaftswissenschaften war ein Kompromiss.

6

Herbert Simon und seine Studenten an der Carnegie-Mellon-Universität legten in den 1980er-Jahren die Grundlagen für unser Verständnis der Expertise. Für eine ausgezeichnete populärwissenschaftliche Einleitung in das Thema vgl. Joshua Foer, *Moonwalk mit Einstein: Wie aus einem vergesslichen Mann ein Gedächtnis-Champion wurde (Moonwalking with Einstein: The Art and Science of Remembering Everything)*, München 2011. Er bezieht sich auf Arbeiten, die fachlich ausführlicher abgehandelt werden in K. Anders Ericsson u.a. (Hg.), *The Cambridge Handbook of Expertise and Expert Performance*, New York 2006.

7

Gary A. Klein, *Natürliche Entscheidungsprozesse. Über die »Quellen der Macht«, die unsere Entscheidungen lenken (Sources of Power)*, Paderborn 2003.

8

Herbert Simon war einer der großen Sozialwissenschaftler des 20. Jahrhunderts, dessen Entdeckungen und Erfindungen von der Politikwissenschaft (wo er seine wissenschaftliche Laufbahn begann) über die Wirtschaftswissenschaften (in denen er mit dem Nobelpreis ausgezeichnet wurde) und die Informatik (zu deren Pionieren er gehörte) bis zur Psychologie reichten.

9

Herbert A. Simon, »What Is an Explanation of Behavior?«, *Psychological Science* 3 (1992): S. 150–161.

10

Das Konzept der Affektheuristik wurde von Paul Slovic entwickelt, Kommilitone von Amos in Michigan und lebenslanger Freund.

11

Vgl. Kapitel 9.

# TEIL I

## 1. Die Figuren der Geschichte

1
Für Darstellungen dieses Zwei-System-Modells vgl. Jonathan St. B. T. Evans und Keith Frankish, (Hg.), *In Two Minds*: *Dual Processes and Beyond*, New York 2009; Jonathan St. B. T. Evans, »Dual-Processing Accounts of Reasoning, Judgment, and Social Cognition«, *Annual Review of Psychology* 59 (2008): S. 255–278. Zu den Pionieren dieses Forschungsgebiets gehören Seymour Epstein, Jonathan Evans, Steven Sloman, Keith Stanovich und Richard West. Ich entlehne die Termini »System 1« und »System 2« aus frühen Arbeiten von Stanovich und West, die einen großen Einfluss auf mein Denken hatten: Keith E. Stanovich und Richard F. West, »Individual Differences in Reasoning: Implications for the Rationality Debate«, *Behavioral and Brain Sciences* 23 (2000): S. 645–665.

2
Dieses subjektive Erleben der Willensfreiheit täuscht manchmal, wie Daniel M. Wegner in *The Illusion of Conscious Will*, Cambridge, MA, 2003, zeigt.

3
Nilli Lavie, »Attention, Distraction and Cognitive Control Under Load«, *Current Directions in Psychological Science* 19 (2010): S. 143–148.

4
In der klassischen Stroop-Aufgabe werden Ihnen verschiedenfarbige Flecken oder verschiedenfarbige Wörter dargeboten. Ihre Aufgabe besteht darin, die Farben zu benennen und die Wörter zu ignorieren. Die Aufgabe ist extrem schwierig, wenn die farbigen Wörter selbst Farbnamen sind (zum Beispiel »Grün« in roter Druckfarbe, gefolgt von »Gelb« in grüner Druckfarbe usw.).

5
Professor Hare schrieb mir: »Ihr Dozent hatte recht.« (16. März 2011). Robert D. Hare, *Gewissenlos. Die Psychopathen unter uns (Without Conscience*: *The Disturbing World of the Psychopaths Among Us)*, Wien, New York 2005. Paul Babiak und Robert D. Hare, *Menschenschinder oder Manager. Psychopathen bei der Arbeit (Snakes in Suits*: *When Psychopaths Go to Work)*, München 2007.

6

Akteure innerhalb des Geistes werden *Homunculi* genannt und sind (weitgehend zu Recht) Objekte des Spotts unter den Wissenschaftlern.
7
Alan D. Baddeley, »Working Memory: Looking Back and Looking Forward«, *Nature Reviews: Neuroscience* 4 (2003): S. 829–838. Alan D. Baddeley, *So denkt der Mensch. Unser Gedächtnis und wie es funktioniert (Your Memory: A User's Guide)* München 1986.

## 2. Aufmerksamkeit und Anstrengung

1
Ein Großteil der Darlegungen in diesem Kapitel stützen sich auf mein Buch *Attention and Effort* (1973). Es kann auf meiner Website (www.princeton.edu/~kahneman/docs/attention_and_effort/Attention_hi_quality.pdf ) kostenlos heruntergeladen werden. Das Hauptthema dieses Buches ist die These einer begrenzten Fähigkeit der Aufmerksamkeitszuwendung und mentalen Anstrengung. Aufmerksamkeit und Anstrengung werden hier als allgemeine Ressourcen behandelt, die für die Ausführung vieler mentaler Aufgaben herangezogen werden können. Die Annahme einer solchen allgemeinen Kapazität ist umstritten, aber sie wurde von anderen Psychologen und Neurowissenschaftlern, die in der Hirnforschung bestätigende Hinweise dafür fanden, erweitert. Vgl. Marcel A. Just und Patricia A. Carpenter, »A Capacity Theory of Comprehension: Individual Differences in Working Memory«, *Psychological Review* 99 (1992): S. 122–149; Marcel A. Just et al., »Neuroindices of Cognitive Workload: Neuroimaging, Pupillometric and Event-Related Potential Studies of Brain Work«, *Theoretical Issues in Ergonomics Science* 4 (2003): S. 56–88. Es gibt auch immer mehr experimentelle Anhaltspunkte für die Existenz universeller Aufmerksamkeitsressourcen, wie sie etwa in Evie Vergauwe et al., »Do Mental Processes Share a Domain-General Resource?«, *Psychological Science* 21 (2010): S. 384–390, beschrieben werden. Und Studien mit bildgebenden Verfahren deuten darauf hin, dass bereits die bloße Erwartung einer anstrengenden Aufgabe, im Vergleich zu einer eher mühelosen gleichartigen Aufgabe, zu einer erhöhten Aktivität in vielen Hirnarealen führt. Carsten N. Boehler et al., »Task-Load-Dependent Activation of Dopaminergic Midbrain Areas in the Absence of Reward«, *Journal of Neuroscience* 31 (2011): S. 4955–4961.

2
Eckhard H. Hess, »Attitude and Pupil Size«, *Scientific American* 212 (1965): S. 46–54.

3
Das Wort »Versuchsperson« (engl. *subject*) klingt in den Ohren mancher Leute zu technisch und reduktionistisch, und die American Psychological Association ermahnt uns, den neutraleren Ausdruck »Teilnehmer« (engl.

*participant*) zu verwenden. Leider ist der politisch korrekte Ausdruck recht lang; das heißt, er beansprucht vergleichsweise viel Platz im Arbeitsgedächtnis und verlangsamt das Denken. Trotzdem werde ich mein Bestes tun, möglichst »Teilnehmer« zu verwenden, aber ich werde zu »Versuchsperson« wechseln, wenn es notwendig erscheint.

4

Daniel Kahneman et al., »Pupillary, Heart Rate, and Skin Resistance Changes During a Mental Task«, *Journal of Experimental Psychology* 79 (1969): S. 164–167.

5

Daniel Kahneman, Jackson Beatty und Irwin Pollack, »Perceptual Deficit During a Mental Task«, *Science* 15 (1967): S. 218–219. Wir verwendeten einen einseitig durchsichtigen Spiegel, sodass die Beobachter die Buchstaben direkt vor sich sahen, während sie in die Kamera blickten. In einer Kontrollgruppe sahen die Teilnehmer den Buchstaben durch eine schmale Blende, um jeglichen Effekt der sich verändernden Pupillengröße auf ihre Sehschärfe auszuschließen. Ihre Erkennungsrate zeigte das umgekehrte V-Muster, das auch bei anderen Probanden beobachtet wurde.

6

Bei dem Versuch, mehrere Aufgaben gleichzeitig auszuführen, können verschiedene Schwierigkeiten auftreten. So ist es zum Beispiel physikalisch unmöglich, zwei verschiedene Dinge genau zur gleichen Zeit zu äußern, und es mag leichter sein, eine akustische und eine visuelle Aufgabe zu kombinieren, als zwei visuelle oder zwei akustische Aufgaben. Bekannte psychologische Theorien versuchen, sämtliche wechselseitigen Interferenzen zwischen Aufgaben auf die Konkurrenz um gesonderte Mechanismen zurückzuführen. Vgl. Alan D. Baddeley, *Working Memory*, New York 1986. Durch Übung mag sich die Fähigkeit zum »Multitasking« in spezifischer Weise verbessern. Doch diese große Vielfalt sehr unterschiedlicher Aufgaben, die sich überlagern, deutet auf die Existenz einer allgemeinen Aufmerksamkeits- oder Anstrengungsressource hin, die zur Bewältigung vieler Aufgaben notwendig ist.

7

Michael E. Smith, Linda K. McEvoy und Alan Gevins, »Neurophysiological Indices of Strategy Development and Skill Acquisition«, *Cognitive Brain Research* 7 (1999): S. 389–404. Alan Gevins et al., »High-Resolution EEG Mapping of Cortical Activation Related to

Working Memory: Effects of Task Difficulty, Type of Processing and Practice«, *Cerebral Cortex* 7 (1997): S. 374–385.

8

So zeigten zum Beispiel Sylvia K. Ahern und Jackson Beatty, dass sich bei Personen, die gute Leistungen im Studierfähigkeitstest erbrachten, die Pupillen bei einer bestimmten Aufgabe weniger stark erweiterten als bei Personen, die bei dem Test schlechter abgeschnitten hatten. »Physiological Signs of Information Processing Vary with Intelligence«, *Science* 205 (1979): S. 1289–1292.

9

Wouter Kool et al., »Decision Making and the Avoidance of Cognitive Demand«, *Journal of Experimental Psychology: General* 139 (2010): S. 665–682. Joseph T. Mc-Guire und Matthew M. Botvinick, »The Impact of Anticipated Demand on Attention and Behavioral Choice«, in: *Effortless Attention*, hg. von Brian Bruya, Cambridge, MA, 2010, S. 103–120.

10

Neurowissenschaftler haben eine Hirnregion identifiziert, die den Gesamtwert einer Handlung bewertet, wenn diese vollständig ausgeführt wurde. Der Anstrengungsaufwand zählt in dieser neuralen Berechnung als ein Kostenfaktor. Joseph T. Mc-Guire und Matthew M. Botvinick, »Prefrontal Cortex, Cognitive Control, and the Registration of Decision Costs«, *PNAS* 107 (2010): S. 7922–7926.

11

Bruno Laeng et al., »Pupillary Stroop Effects«, *Cognitive Processing* 12 (2011): S. 13–21.

12

Michael I. Posner und Mary K. Rothbart, »Research on Attention Networks as a Model for the Integration of Psychological Science«, *Annual Review of Psychology* 58 (2007): S. 1–23. John Duncan et al., »A Neural Basis for General Intelligence«, *Science* 289 (2000): S. 457–460.

13

Stephen Monsell, »Task Switching«, *Trends in Cognitive Sciences* 7 (2003): S. 134–14 0.

14

Baddeley, *Working Memory.*

15

Andrew A. Conway, Michael J. Kane und Randall W. Engle, »Working Memory Capacity and Its Relation to General Intelligence«, *Trends in Cognitive Sciences* 7 (2003): S. 547–552.

16

Daniel Kahneman, Rachel Ben-Ishai und Michael Lotan, »Relation of a Test of Attention to Road Accidents«, *Journal of Applied Psychology* 58 (1973): S. 113–115. Daniel Gopher, »A Selective Attention Test as a Predictor of Success in Flight Training«, *Human Factors* 24 (1982): S. 173–183.

## 3. Der faule Kontrolleur

1
Mihaly Csikszentmihalyi, *Die außergewöhnliche Erfahrung im Alltag. Die Psychologie des Flow-Erlebnisses (Flow: The Psychology of Optimal Experience)*, Stuttgart 1991).
2
Baba Shiv und Alexander Fedorikhin, »Heart and Mind in Conflict: The Interplay of Affect and Cognition in Consumer Decision Making«, *Journal of Consumer Research* 26 (1999): S. 278–292. Malte Friese, Wilhelm Hofmann und Michaela Wänke, »When Impulses Take Over: Moderated Predictive Validity of Implicit and Explicit Attitude Measures in Predicting Food Choice and Consumption Behaviour«, *British Journal of Social Psychology* 47 (2008): S. 397–419.
3
Daniel T. Gilbert, »How Mental Systems Believe«, *American Psychologist* 46 (1991): S. 107–119. C. Neil Macrae und Galen V. Bodenhausen, »Social Cognition: Thinking Categorically about Others«, *Annual Review of Psychology* 51 (2000): S. 93–120.
4
Sian L. Beilock und Thomas H. Carr, »When High-Powered People Fail: Working Memory and Choking Under Pressure in Math«, *Psychological Science* 16 (2005): S. 101–105.
5
Martin S. Hagger et al., »Ego Depletion and the Strength Model of Self-Control: A Meta-Analysis«, *Psychological Bulletin* 136 (2010): S. 495–525.
6
Mark Muraven und Elisaveta Slessareva, »Mechanisms of Self-Control Failure: Motivation and Limited Resources«, *Personality and Social Psychology Bulletin* 29 (2003): S. 894–906. Mark Muraven, Dianne M. Tice und Roy F. Baumeister, »Self-Control as a Limited Resource: Regulatory Depletion Patterns«, *Journal of Personality and Social Psychology* 74 (1998): S. 774–89.
7
Matthew T. Gailliot et al., »Self-Control Relies on Glucose as a Limited Energy Source: Willpower Is More Than a Metaphor«, *Journal of*

*Personality and Social Psychology* 92 (2007): S. 325–336. Matthew T. Gailliot und Roy F. Baumeister, »The Physiology of Willpower: Linking Blood Glucose to Self-Control«, *Personality and Social Psychology Review* 11 (2007): S. 303–327.

8

Gailliot, »Self-Control Relies on Glucose as a Limited Energy Source«.

9

Shai Danziger, Jonathan Levav und Liora Avnaim-Pesso, »Extraneous Factors in Judicial Decisions«, *PNAS* 108 (2011): S. 6889–6892.

10

Shane Frederick, »Cognitive Reflection and Decision Making«, *Journal of Economic Perspectives* 19 (2005): S. 25–42.

11

Dieser systematische Fehler wird *belief bias* (»Überzeugungsfehler«) genannt. Evans, »Dual-Processing Accounts of Reasoning, Judgment, and Social Cognition«.

12

Keith E. Stanovich, *Rationality and the Reflective Mind*, New York 2011.

13

Walter Mischel und Ebbe B. Ebbesen, »Attention in Delay of Gratification«, *Journal of Personality and Social Psychology* 16 (1970): S. 329–337.

14

Inge-Marie Eigsti et al., »Predicting Cognitive Control from Preschool to Late Adolescence and Young Adulthood«, *Psychological Science* 17 (2006): S. 478–484.

15

Mischel und Ebbesen, »Attention in Delay of Gratification«. Walter Mischel, »Processes in Delay of Gratification«, in *Advances in Experimental Social Psychology*, Vol. 7, hg. von Leonard Berkowitz, San Diego 1974, *S.* 249–292. Walter Mischel, Yuichi Shoda und Monica L. Rodriguez, »Delay of Gratification in Children«, *Science* 244 (1989): S. 933–938. Eigsti, »Predicting Cognitive Control from Preschool to Late Adolescence«.

16

M. Rosario Rueda et al., »Training, Maturation, and Genetic Influences on the Development of Executive Attention«, *PNAS* 102 (2005): S. 14931–

14936.

17

Maggie E. Toplak, Richard F. West und Keith E. Stanovich, »The Cognitive Reflection Test as a Predictor of Performance on Heuristics-and-Biases Tasks«, *Memory & Cognition* (im Druck).

## 4. Die Assoziationsmaschine

1
Carey K. Morewedge und Daniel Kahneman, »Associative Processes in Intuitive Judgment«, *Trends in Cognitive Sciences* 14 (2010): S. 435–440.
2
Um keine Verwirrung zu stiften, habe ich im Text nicht erwähnt, dass sich auch Ihre Pupille erweiterte. Die Pupille erweitert sich sowohl bei emotionalem *Arousal* als auch bei einem *Arousal*, das mit intellektueller Anstrengung einhergeht.
3
Paula M. Niedenthal, »Embodying Emotion«, *Science* 316 (2007): S. 1002–1005.
4
Das Bild ist der Funktionsweise einer Pumpe entlehnt. Die ersten Pumpbewegungen fördern noch keine Flüssigkeit hervor, aber sie ermöglichen es nachfolgenden Pumpbewegungen, effektiv zu sein.
5
John A. Bargh, Mark Chen und Lara Burrows, »Automaticity of Social Behavior: Direct Effects of Trait Construct and Stereotype Activation on Action«, *Journal of Personality and Social Psychology* 71 (1996): S. 230–244.
6
Thomas Mussweiler, »Doing Is for Thinking! Stereotype Activation by Stereotypic Movements«, *Psychological Science* 17 (2006): S. 17–21.
7
Fritz Strack, Leonard L. Martin und Sabine Stepper, »Inhibiting and Facilitating Conditions of the Human Smile: A Nonobtrusive Test of the Facial Feedback Hypothesis«, *Journal of Personality and Social Psychology* 54 (1988): S. 768–777.
8
Ulf Dimberg, Monika Thunberg und Sara Grunedal, »Facial Reactions to Emotional Stimuli: Automatically Controlled Emotional Responses«, *Cognition and Emotion* 16 (2002): S. 449–471.
9

Gary L. Wells und Richard E. Petty, »The Effects of Overt Head Movements on Persuasion: Compatibility and Incompatibility of Responses«, *Basic and Applied Social Psychology* 1 (1980): S. 219–230.

10

Jonah Berger, Marc Meredith und S. Christian Wheeler, »Contextual Priming: Where People Vote Affects How They Vote«, *PNAS* 105 (2008): S. 8846–8849.

11

Kathleen D. Vohs, »The Psychological Consequences of Money«, *Science* 314 (2006): S. 1154–1156.

12

Jeff Greenberg et al., »Evidence for Terror Management Theory II: The Effect of Mortality Salience on Reactions to Those Who Threaten or Bolster the Cultural Worldview«, *Journal of Personality and Social Psychology* 58 (1990): S. 308–318.

13

Chen-Bo Zhong und Katie Liljenquist, »Washing Away Your Sins: Threatened Morality and Physical Cleansing«, *Science* 313 (2006): S. 1451–1452.

14

Spike Lee und Norbert Schwarz, »Dirty Hands and Dirty Mouths: Embodiment of the Moral-Purity Metaphor Is Specific to the Motor Modality Involved in Moral Transgression«, *Psychological Science* 21 (2010): S. 1423–1425.

15

Melissa Bateson, Daniel Nettle und Gilbert Roberts, »Cues of Being Watched Enhance Cooperation in a Real-World Setting«, *Biology Letters* 2 (2006): S. 412–414.

16

Timothy Wilson führt in seinem Buch *Gestatten, mein Name ist Ich. Das adaptive Unbewusste – eine psychologische Entdeckungsreise (Strangers to Ourselves)*, München, Zürich 2007, das Konzept eines »adaptiven Unbewussten« ein, das große Ähnlichkeit mit System 1 besitzt.

## 5. Kognitive Leichtigkeit

1
Der Fachbegriff für kognitive Leichtigkeit ist *fluency* (»Verarbeitungsflüssigkeit«).
2
Adam L. Alter und Daniel M. Oppenheimer, »Uniting the Tribes of Fluency to Form a Metacognitive Nation«, *Personality and Social Psychology Review* 13 (2009): S. 219–235.
3
Larry L. Jacoby, Colleen Kelley, Judith Brown und Jennifer Jasechko, »Becoming Famous Overnight: Limits on the Ability to Avoid Unconscious Influences of the Past«, *Journal of Personality and Social Psychology* 56 (1989): S. 326–338.
4
Bruce W. A. Whittlesea, Larry L. Jacoby und Krista Girard, »Illusions of Immediate Memory: Evidence of an Attributional Basis for Feelings of Familiarity and Perceptual Quality«, *Journal of Memory and Language* 29 (1990): S. 716–732.
5
Wenn Sie einem Freund begegnen, können Sie ihn normalerweise sofort zuordnen und benennen; Sie wissen meistens, wann Sie ihm zum letzten Mal begegnet sind, was er trug und was Sie zueinander sagten. Das Gefühl der Vertrautheit wird nur dann relevant, wenn solche spezifischen Erinnerungen nicht verfügbar sind. Es ist ein Notbehelf. Obwohl dieser Notbehelf nicht hundertprozentig verlässlich ist, ist er doch viel besser als nichts. Es ist das Gefühl der Vertrautheit, das Sie vor der Verlegenheit bewahrt, überrascht zu sein (und zu tun), wenn Sie von jemandem, der Ihnen nur vage vertraut vorkommt, als ein alter Freund begrüßt werden.
6
Ian Begg, Victoria Armour und Thérèse Kerr, »On Believing What We Remember«, *Canadian Journal of Behavioural Science* 17 (1985): S. 199–214.
7
Daniel M. Oppenheimer, »Consequences of Erudite Vernacular Utilized Irrespective of Necessity: Problems with Using Long Words Needlessly«,

*Applied Cognitive Psychology* 20 (2006): S. 139–156.

8

Matthew S. McGlone und Jessica Tofighbakhsh, »Birds of a Feather Flock Conjointly (?): Rhyme as Reason in Aphorisms«, *Psychological Science* 11 (2000): S. 424–428.

9

Anuj K. Shah und Daniel M. Oppenheimer, »Easy Does It: The Role of Fluency in Cue Weighting«, *Judgment and Decision Making Journal* 2 (2007): S. 371–379.

10

Adam L. Alter, Daniel M. Oppenheimer, Nicholas Epley und Rebecca Eyre, »Overcoming Intuition: Metacognitive Difficulty Activates Analytic Reasoning«, *Journal of Experimental Psychology: General* 136 (2007): S. 569–576.

11

5, 47

12

Piotr Winkielman und John T. Cacioppo, »Mind at Ease Puts a Smile on the Face: Psychophysiological Evidence That Processing Facilitation Increases Positive Affect«, *Journal of Personality and Social Psychology* 81 (2001): S. 989–1000.

13

Adam L. Alter und Daniel M. Oppenheimer, »Predicting Short-Term Stock Fluctuations by Using Processing Fluency«, *PNAS* 103 (2006). Michael J. Cooper, Orlin Dimitrov und P. Raghavendra Rau, »A Rose.com by Any Other Name«, *Journal of Finance* 56 (2001): S. 2371–2388.

14

Pascal Pensa, »Nomen Est Omen: How Company Names Influence Short- and Long-Run Stock Market Performance«, *Social Science Research Network Working Paper*, September 2006.

15

Robert B. Zajonc, »Attitudinal Effects of Mere Exposure«, *Journal of Personality and Social Psychology* 9 (1968): S. 1–27.

16

Robert B. Zajonc und D. W. Rajecki, »Exposure and Affect: A Field Experiment«, *Psychonomic Science* 17 (1969): S. 216–217.

17

Jennifer L. Monahan, Sheila T. Murphy und Robert B. Zajonc, »Subliminal Mere Exposure: Specific, General, and Diffuse Effects«, *Psychological Science* 11 (2000): S. 462–466.

18

D. W. Rajecki, »Effects of Prenatal Exposure to Auditory or Visual Stimulation on Postnatal Distress Vocalizations in Chicks«, *Behavioral Biology* 11 (1974): S. 525–536.

19

Robert B. Zajonc, »Mere Exposure: A Gateway to the Subliminal«, *Current Directions in Psychological Science* 10 (2001): S. 227.

20

Annette Bolte, Thomas Goschke und Julius Kuhl, »Emotion and Intuition: Effects of Positive and Negative Mood on Implicit Judgments of Semantic Coherence«, *Psychological Science* 14 (2003): S. 416–421.

21

Die Analyse schließt sämtliche Fälle aus, in denen die Versuchsperson die richtige Lösung fand. Sie zeigt, dass selbst Probanden, die letztlich keine gemeinsame Assoziation finden, eine Ahnung davon haben, ob eine gefunden werden kann.

22

Sascha Topolinski und Fritz Strack, »The Architecture of Intuition: Fluency and Affect Determine Intuitive Judgments of Semantic and Visual Coherence and Judgments of Grammaticality in Artificial Grammar Learning«, *Journal of Experimental Psychology: General* 138 (2009): S. 39–63.

23

Bolte, Goschke und Kuhl, »Emotion and Intuition«.

24

Barbara Fredrickson, *Die Macht der guten Gefühle. Wie eine positive Haltung Ihr Leben dauerhaft verändert (Positivity: Groundbreaking Research Reveals How to Embrace the Hidden Strength of Positive Emotions, Overcome Negativity, and Thrive)*, Franfurt a. M., New York 2011. Joseph P. Forgas und Rebekah East, »On Being Happy and Gullible: Mood Effects on Skepticism and the Detection of Deception«, *Journal of Experimental Social Psychology* 44 (2008): S. 1362–1367.

25

Sascha Topolinski et al., »The Face of Fluency: Semantic Coherence Automatically Elicits a Specific Pattern of Facial Muscle Reactions«, *Cognition and Emotion* 23 (2009): S. 260–271.

26

Sascha Topolinski und Fritz Strack, »The Analysis of Intuition: Processing Fluency and Affect in Judgments of Semantic Coherence«, *Cognition and Emotion* 23 (2009): S. 1465–1503.

## 6. Normen, Überraschungen und Ursachen

1
Daniel Kahneman und Dale T. Miller, »Norm Theory: Comparing Reality to Its Alternatives«, *Psychological Review* 93 (1986): S. 136–153.
2
Jos J. A. Van Berkum, »Understanding Sentences in Context: What Brain Waves Can Tell Us«, *Current Directions in Psychological Science* 17 (2008): 376–380.
3
Ran R. Hassin, John A. Bargh und James S. Uleman, »Spontaneous Causal Inferences«, *Journal of Experimental Social Psychology* 38 (2002): S. 515–522.
4
Albert Michotte, *The Perception of Causality*, Andover 1963). Alan M. Leslie und Stephanie Keeble, »Do Six-Month-Old Infants Perceive Causality?«, *Cognition* 25 (1987): S. 265–288.
5
Fritz Heider und Marianne Simmel, »An Experimental Study of Apparent Behavior«, *American Journal of Psychology* 13 (1944): S. 243–259.
6
Leslie und Keeble, »Do Six-Month-Old Infants Perceive Causality?«.
7
Paul Bloom, »Is God an Accident?«, *Atlantic*, December 2005.

## 7. Eine Maschine für voreilige Schlussfolgerungen

1
Daniel T. Gilbert, Douglas S. Krull und Patrick S. Malone, »Unbelieving the Unbelievable: Some Problems in the Rejection of False Information«, *Journal of Personality and Social Psychology* 59 (1990): S. 601–613.

2
Solomon E. Asch, »Forming Impressions of Personality«, *Journal of Abnormal and Social Psycholo*gy 41 (1946): S. 258–290.

3
Ebd.

4
James Surowiecki, *Die Weisheit der Vielen. Warum Gruppen klüger sind als Einzelne und wie wir das kollektive Wissen für unser wirtschaftliches, soziales und politisches Handeln nutzen können (The Wisdom of Crowds)*, München 2005.

5
Lyle A. Brenner, Derek J. Koehler und Amos Tversky, »On the Evaluation of One-Sided Evidence«, *Journal of Behavioral Decision Making* 9 (1996): S. 59–70.

## 8. Wie wir Urteile bilden

1
Alexander Todorov, Sean G. Baron und Nikolaas N. Oosterhof, »Evaluating Face Trustworthiness: A Model-Based Approach«, *Social Cognitive and Affective Neurocience* 3 (2008): S. 119–127.
2
Alexander Todorov, Chris P. Said, Andrew D. Engell und Nikolaas N. Oosterhof, »Understanding Evaluation of Faces on Social Dimensions«, *Trends in Cognitive Sciences* 12 (2008): S. 455–460.
3
Alexander Todorov, Manish Pakrashi und Nikolaas N. Oosterhof, »Evaluating Faces on Trustworthiness After Minimal Time Exposure«, *Social Cognition* 27 (2009): S. 813–833.
4
Alexander Todorov et al., »Inference of Competence from Faces Predict Election Outcomes«, *Science* 308 (2005): S. 1623–1626. Charles C. Ballew und Alexander Todorov, »Predicting Political Elections from Rapid and Unreflective Face Judgments«, *PNAS* 104 (2007): S. 17948–17953. Christopher Y. Olivola und Alexander Todorov, »Elected in 100 Milliseconds: Appearance-Based Trait Inferences and Voting«, *Journal of Nonverbal Behavior* 34 (2010): S. 83–110.
5
Gabriel Lenz und Chappell Lawson, »Looking the Part: Television Leads Less Informed Citizens to Vote Based on Candidates' Appearance«, *American Journal of Political Science* 55 (2011): S. 574–589.
6
Amos Tversky und Daniel Kahneman, »Extensional Versus Intuitive Reasoning: The Conjunction Fallacy in Probability Judgment«, *Psychological Review* 90 (1983): S. 293–315.
7
William H. Desvousges et al., »Measuring Natural Resource Damages with Contingent Valuation: Tests of Validity and Reliability«, in *Contingent Valuation: A Critical Assessment*, hg. von Jerry A. Hausman, Amsterdam 1993, S. 91–159.
8

Stanley S. Stevens, *Psychophysics*: *Introduction to Its Perceptual, Neural, and Social Prospect*, New York 1975.

9

Mark S. Seidenberg und Michael K. Tanenhaus, »Orthographic Effects on Rhyme Monitoring«, *Journal of Experimental Psychology: Human Learning and Memory* 5 (1979): S. 546–554.

10

Sam Glucksberg, Patricia Gildea und Howard G. Bookin, »On Understanding Nonliteral Speech: Can People Ignore Metaphors?«, *Journal of Verbal Learning and Verbal Behavior* 21 (1982): S. 85–98.

## 9. Eine leichtere Frage beantworten

[1]
Eine alternative Theorie der Urteilsheuristiken wurde von Gerd Gigerenzer, Peter M. Todd und der ABC Research Group vorgestellt in *Simple Heuristics That Make Us Smart*, New York 1999. Sie beschreiben »schnelle und sparsame« formale Verfahren wie etwa die »Take-the-best-[cue]«-Heuristik (»Nehmen Sie den besten Hinweis«), die unter bestimmten Umständen auf der Basis geringer Information recht zutreffende Urteile generiert. Gigerenzer hat betont, dass sich seine Heuristiken von denen, die Amos und ich erforscht haben, unterscheiden, und er hatte gegenüber den Verzerrungen, zu denen sie zwangsläufig führen, vor allem ihre Genauigkeit hervorgehoben. Ein Großteil der Forschungen, die den Nutzen schneller und sparsamer Heuristiken belegen, basieren auf statistischen Simulationen, die zeigen, dass sie in bestimmten realen Situationen hilfreich sein *könnten*. Aber die empirischen Belege für die tatsächliche psychologische Relevanz dieser Heuristiken sind nach wie vor dürftig und umstritten. Die bemerkenswerteste Entdeckung, die mit diesem Ansatz verbunden ist, ist die Rekognitionsheuristik, die durch ein mittlerweile recht bekanntes Beispiel veranschaulicht wird: Eine Versuchsperson, die gefragt wird, welche von zwei Städten größer ist, und die eine der beiden Städte wiedererkennt, nimmt im Allgemeinen an, dass die ihr bekannte Stadt die größere ist. Die Rekognitionsheuristik funktioniert recht gut, wenn die Versuchsperson weiß, dass die Stadt, die sie wiedererkennt, groß ist; weiß sie indes, dass sie klein ist, wird sie vernünftigerweise annehmen, dass die unbekannte Stadt größer ist. Anders als die Theorie behauptet, benutzen die Probanden mehr als nur den Rekognitionshinweis(reiz): Daniel M. Oppenheimer, »Not So Fast! (and Not So Frugal!): Rethinking the Recognition Heuristic«, *Cognition* 90 (2003): B1 – B9. Eine Schwäche der Theorie besteht darin, dass Heuristiken nach allem, was wir über kognitive Prozesse wissen, nicht sparsam sein müssen. Das Gehirn verarbeitet parallel riesige Informationsmengen, und es kann schnell zutreffende Urteile fällen, ohne Informationen zu ignorieren. Außerdem wissen wir seit den ersten Studien über Schachmeister, dass Fertigkeit nicht unbedingt darin bestehen muss, dass man lernt, weniger Informationen zu nutzen. Im Gegenteil,

Fertigkeit ist häufiger die Fähigkeit, große Informationsmengen schnell und effizient zu verarbeiten.

2

Fritz Strack, Leonard L. Martin und Norbert Schwarz, »Priming and Communication: Social Determinants of Information Use in Judgments of Life Satisfaction«, *European Journal of Social Psychology* 18 (1988): S. 429–442.

3

Die Korrelation betrug 0,66.

4

Weitere Ersetzungsthemen sind Zufriedenheit in der Ehe, Zufriedenheit im Beruf und Zufriedenheit in der Freizeit: Norbert Schwarz, Fritz Strack und Hans-Peter Mai, »Assimilation and Contrast Effects in Part-Whole Question Sequences: A Conversational Logic Analysis«, *Public Opinion Quarterly* 55 (1991): S. 3–23.

5

Eine in Deutschland durchgeführte telefonische Umfrage beinhaltete eine Frage über die allgemeine Lebenszufriedenheit. Als die Selbsteinschätzungen der Zufriedenheit mit dem lokalen Wetter zum Zeitpunkt der Befragung in Beziehung gesetzt wurden, zeigte sich eine hohe Korrelation. Bekanntlich schwankt die Stimmung mit dem Wetter, und die Ersetzung erklärt den Effekt auf die angegebene Zufriedenheit. Eine andere Version der telefonischen Befragung ergab jedoch ein etwas anderes Ergebnis. Diese Personen wurden nach dem aktuellen Wetter gefragt, bevor ihnen die Frage nach der Zufriedenheit gestellt wurde. Bei ihnen wirkte sich das Wetter in keiner Weise auf die berichtete Zufriedenheit aus! Das ausdrückliche Wetter-Priming lieferte ihnen eine Erklärung für ihre Stimmung und unterminierte die Verknüpfung, die normalerweise zwischen der aktuellen Stimmung und der allgemeinen Lebenszufriedenheit hergestellt wird.

6

Melissa L. Finucane et al., »The Affect Heuristic in Judgments of Risks and Benefits«, *Journal of Behavioral Decision Making* 13 (2000): S. 1–17.

# TEIL II

## 10. Das Gesetz der kleinen Zahlen

1
Howard Wainer and Harris L. Zwerling, »Evidence That Smaller Schools Do Not Improve Student Achievement«, *Phi Delta Kappan* 88 (2006): S. 300–303. Das Beispiel wurde diskutiert von Andrew Gelman und Deborah Nolan, *Teaching Statistics: A Bag of Tricks*, New York 2002.

2
Jacob Cohen, »The Statistical Power of Abnormal-Social Psychological Research: A Review«, *Journal of Abnormal and Social Psychology* 65 (1962): S. 145–153.

3
Amos Tversky und Daniel Kahneman, »Belief in the Law of Small Numbers«, *Psychological Bulletin* 76 (1971): S. 105–110.

4
Der Gegensatz, den wir zwischen Intuition und Berechnung aufstellten, scheint auf die Unterscheidung zwischen System 1 und System 2 hinzudeuten, aber wir waren damals noch weit entfernt von der Perspektive dieses Buches. Wir verwendeten das Wort »Intuition« für alles, was keine Berechnung ist, für jedes informelle Verfahren, um zu einer Schlussfolgerung zu gelangen.

5
William Feller, *Introduction to Probability Theory and Its Applications*, New York 1950.

6
Thomas Gilovich, Robert Vallone und Amos Tversky, »The Hot Hand in Basketball: On the Misperception of Random Sequences«, *Cognitive Psychology* 17 (1985): S. 295–314.

## 11. Anker

1
Robyn LeBoeuf und Eldar Shafir, »The Long and Short of It: Physical Anchoring Effects«, *Journal of Behavioral Decision Making* 19 (2006): S. 393–406.
2
Nicholas Epley und Thomas Gilovich, »Putting Adjustment Back in the Anchoring and Adjustment Heuristic: Differential Processing of Self-Generated and Experimenter-Provided Anchors«, *Psychological Science* 12 (2001): S. 391–396.
3
Epley und Gilovich, »The Anchoring-and-Adjustment Heuristic«.
4
Thomas Mussweiler, »The Use of Category and Exemplar Knowledge in the Solution of Anchoring Tasks«, *Journal of Personality and Social Psychology* 78 (2000): S. 1038–1052.
5
Karen E. Jacowitz und Daniel Kahneman, »Measures of Anchoring in Estimation Tasks«, *Personality and Social Psychology Bulletin* 21 (1995): S. 1161–1166.
6
Gregory B. Northcraft und Margaret A. Neale, »Experts, Amateurs, and Real Estate: An Anchoring-and-Adjustment Perspective on Property Pricing Decisions«, *Organizational Behavior and Human Decision Processes* 39 (1987): S. 84–97. Der hohe Anker lag 12 Prozent über dem Listenpreis, der niedrige 12 Prozent unter diesem Preis.
7
Birte Englich, Thomas Mussweiler und Fritz Strack, »Playing Dice with Criminal Sentences: The Influence of Irrelevant Anchors on Experts' Judicial Decision Making«, *Personality and Social Psychology Bulletin* 32 (2006): S. 188–200.
8
Brian Wansink, Robert J. Kent und Stephen J. Hoch, »An Anchoring and Adjustment Model of Purchase Quantity Decisions«, *Journal of Marketing Research* 35 (1998): S. 71–81.

9
Adam D. Galinsky und Thomas Mussweiler, »First Offers as Anchors: The Role of Perspective-Taking and Negotiator Focus«, *Journal of Personality and Social Psychology* 81 (2001): S. 657–669.
10
Greg Pogarsky und Linda Babcock, »Damage Caps, Motivated Anchoring, and Bargaining Impasse«, *Journal of Legal Studies* 30 (2001): S. 143–159.
11
Für eine experimentelle Demonstration vgl. Chris Guthrie, Jeffrey J. Rachlinski und Andrew J. Wistrich, »Judging by Heuristic-Cognitive Illusions in Judicial Decision Making«, *Judicature* 86 (2002): S. 44–50.

## 12. Die Wissenschaft der Verfügbarkeit

1
Amos Tversky und Daniel Kahneman, »Availability: A Heuristic for Judging Frequency and Probability«, *Cognitive Psychology* 5 (1973): S. 207–232.
2
Michael Ross und Fiore Sicoly, »Egocentric Biases in Availability and Attribution«, *Journal of Personality and Social Psychology* 37 (1979): S. 322–336.
3
Schwarz et al., »Ease of Retrieval as Information«.
4
Sabine Stepper und Fritz Strack, »Proprioceptive Determinants of Emotional and Nonemotional Feelings«, *Journal of Personality and Social Psychology* 64 (1993): S. 211–220.
5
Für eine Zusammenfassung dieses Forschungsfelds vgl. Rainer Greifeneder, Herbert Bless und Michel T. Pham, »When Do People Rely on Affective and Cognitive Feelings in Judgment? A Review«, *Personality and Social Psychology Review* 15 (2011): S. 107–141.
6
Alexander Rotliman und Norbert Schwarz, »Constructing Perceptions of Vulnerability: Personal Relevance and the Use of Experimental Information in Health Judgments«, *Personality and Social Psychology Bulletin* 24 (1998): S. 1053–1064.
7
Rainer Greifeneder und Herbert Bless, »Relying on Accessible Content Versus Accessibility Experiences: The Case of Processing Capacity«, *Social Cognition* 25 (2007): S. 853–881.
8
Markus Ruder und Herbert Bless, »Mood and the Reliance on the Ease of Retrieval Heuristic«, *Journal of Personality and Social Psychology* 85 (2003): S. 20–32.
9

Rainer Greifeneder und Herbert Bless, »Depression and Reliance on Ease-of-Retrieval Experiences«, *European Journal of Social Psychology* 38 (2008): S. 213–230.

10

Chezy Ofir et al., »Memory-Based Store Price Judgments: The Role of Knowledge and Shopping Experience«, *Journal of Retailing* 84 (2008): S. 414–423.

11

Eugene M. Caruso, »Use of Experienced Retrieval Ease in Self and Social Judgments«, *Journal of Experimental Social Psychology* 44 (2008): S. 148–155.

12

Johannes Keller und Herbert Bless, »Predicting Future Affective States: How Ease of Retrieval and Faith in Intuition Moderate the Impact of Activated Content«, *European Journal of Social Psychology* 38 (2008): S. 1–10.

13

Mario Weick und Ana Guinote, »When Subjective Experiences Matter: Power Increases Reliance on the Ease of Retrieval«, *Journal of Personality and Social Psychology* 94 (2008): S. 956–970.

## 13. Verfügbarkeit, Emotion und Risiko

1
Damasios Konzept, die sogenannte Hypothese der somatischen Marker, ist auf breite Unterstützung gestoßen: Antonio R. Damasio, *Descartes' Irrtum: Fühlen, Denken und das menschliche Gehirn*, München 1994. Antonio R. Damasio, »The Somatic Marker Hypothesis and the Possible Functions of the Prefrontal Cortex«, *Philosophical Transactions: Biological Sciences* 351 (1996): S. 1413–1420.

2
Finucane et al., »The Affect Heuristic in Judgments of Risks and Benefits«. Paul Slovic, Melissa Finucane, Ellen Peters und Donald G. MacGregor, »The Affect Heuristic«, in Thomas Gilovich, Dale Griffin und Daniel Kahneman, Hg., *Heuristics and Biases*, New York 2002, S. 397–420. Paul Slovic, Melissa Finucane, Ellen Peters und Donald G. MacGregor, »Risk as Analysis and Risk as Feelings: Some Thoughts About Affect, Reason, Risk, and Rationality«, *Risk Analysis* 24 (2004): S. 1–12. Paul Slovic, »Trust, Emotion, Sex, Politics, and Science: Surveying the Risk-Assessment Battlefield«, *Risk Analysis* 19 (1999): S. 689–701.

3
Slovic, »Trust, Emotion, Sex, Politics, and Science«. Die in dieser Studie verwendeten Technologien sind keine alternativen Lösungen für dasselbe Problem. Bei realistischen Problemen, bei denen konkurrierende Lösungen in Betracht gezogen werden, muss die Korrelation zwischen Kosten und Nutzen negativ sein; die Lösungen, die den größten Nutzen bringen, sind zugleich die kostspieligsten. Es ist eine interessante Frage, ob Laien und auch Experten selbst in diesen Fällen den richtigen Zusammenhang nicht erkennen.

4
Jonathan Haidt, »The Emotional Dog and Its Rational Tail: A Social Institutionist Approach to Moral Judgment«, *Psychological Review* 108 (2001): S. 814–834.

5
Paul Slovic, *The Perception of Risk*, Sterling 2000.

6

Timur Kuran und Cass R. Sunstein, »Availability Cascades and Risk Regulation«, *Stanford Law Review* 51 (1999): S. 683–768. *CERCLA*, the Comprehensive Environmental Response, Compensation, and Liability Act, 1980 ratifiziert.

7

Paul Slovic, der im Alar-Fall als Experte für die Apfelanbauer auftrat, hat eine ganze andere Sichtweise: »Der Schrecken wurde ausgelöst durch die CBS-Sendung *60 Minutes*, in der berichtet wurde, dass 4000 Kinder an Krebs sterben würden (ohne dass hier genauere Angaben über die Erkrankungsrisiken gemacht wurden), und diese Nachricht wurde begleitet von erschreckenden Bildern kahlköpfiger Kinder auf einer onkologischen Station – und vielen weiteren falschen Behauptungen. Die Geschichte enthüllte auch die Inkompetenz der US-Umweltschutzbehörde bei der Bewertung der Sicherheit von Alar und ihr mangelhaftes Krisenmanagement, was das Vertrauen in die Zuverlässigkeit der Aufsichtsbehörde zerstörte. In Anbetracht dessen halte ich die Reaktion der Öffentlichkeit für rational.« (Persönliche Mitteilung, 11. Mai 2011)

## 14. Was studiert Tom W.?

1
Dieses Beispiel stammt von Max H. Bazerman und Don A. Moore, *Judgment in Managerial Decision Making*, New York 2008.
2
Jonathan St. B. T. Evans, »Heuristic and Analytic Processes in Reasoning«, *British Journal of Psychology* 75 (1984): S. 451–468.
3
Norbert Schwarz et al., »Base Rates, Representativeness, and the Logic of Conversation: The Contextual Relevance of ›Irrelevant‹ Information«, *Social Cognition* 9 (1991): S. 67–84.
4
Alter, Oppenheimer, Epley und Eyre, »Overcoming Intuition«.
5
Die einfachste Form der Bayesschen Regel ist die Odds-Form: A-posteriori-Odds = A-priori-Odds × Wahrscheinlichkeitsverhältnis, wobei die A-posteriori-Odds die Odds (das Verhältnis der Wahrscheinlichkeiten) für zwei konkurrierende Hypothesen sind. Nehmen wir ein Problem der Diagnose. Ihr Freund wurde positiv auf eine schwerwiegende Erkrankung getestet. Diese Erkrankung ist selten: Nur einer von 600 Fällen, die getestet werden, hat diese Erkrankung. Der Test ist ziemlich genau. Sein Wahrscheinlichkeitsquotient liegt bei 25: 1, was bedeutet, dass die Wahrscheinlichkeit, dass eine Person, die die Krankheit hat, ein positives Testergebnis aufweist, 25-mal größer ist als die Wahrscheinlichkeit eines falsch-positiven Testergebnisses. Ein positives Testergebnis ist eine erschreckende Nachricht, aber die Odds, dass Ihr Freund die Krankheit hat, ist nur von 1/600 auf 25/600 angestiegen, und die Wahrscheinlichkeit beträgt 4 Prozent. Für die Hypothese, dass Tom W. Informatiker ist, betragen die A-priori-Odds, die einer Basisrate von 3 Prozent entsprechen, 0,03/0,97 = 0,031. Bei einem Wahrscheinlichkeitsverhältnis von vier (die Wahrscheinlichkeit, dass die Beschreibung zutrifft, ist viermal höher, wenn Tom W. Informatiker ist, als wenn er kein Informatiker ist) betragen die A-posteriori-Odds 4 × 0,031 = 12,4. Aus diesen Odds kann man berechnen, dass die A-posteriori-Wahrscheinlichkeit, dass Tom W. Informatiker ist, jetzt 11 Prozent beträgt (weil 12,4/112,4 = 0,11).

## 15. Linda: Weniger ist mehr

1
Amos Tversky und Daniel Kahneman, »Extensional Versus Intuitive Reasoning: The Conjunction Fallacy in Probability Judgment«, *Psychological Review* 90 (1983), S. 293–315.

2
Vgl. Stephen Jay Gould, *Bravo, Brontosaurus. Die verschlungenen Wege der Naturgeschichte (Bully for Brontosaurus)*, Hamburg 1994.

3
Vgl. u. a. Ralph Hertwig und Gerd Gigerenzer, »The ›Conjunction Fallacy‹ Revisited: How Intelligent Inferences Look Like Reasoning Errors«, *Journal of Behavioral Decision Making* 12 (1999): S. 275–305; Ralph Hertwig, Bjoern Benz und Stefan Krauss, »The Conjunction Fallacy and the Many Meanings of And«, *Cognition* 108 (2008): S. 740–753.

4
Barbara Mellers, Ralph Hertwig und Daniel Kahneman, »Do Frequency Representations Eliminate Conjunction Effects? An Exercise in Adversarial Collaboration«, *Psychological Science* 12 (2001): S. 269–275.

## 16. Ursachen vs. Statistik

1
Bei Anwendung der Bayesschen Regel in Odds-Form gilt, dass die A-priori-Odds die Odds für das Blue-Taxi aus der Basisrate sind, und das Wahrscheinlichkeitsverhältnis ist die Wahrscheinlichkeit, dass der Zeuge aussagt, das Taxi sei blau, wenn es (tatsächlich) blau ist, dividiert durch die Wahrscheinlichkeit, dass der Zeuge aussagt, das Taxi sei blau, während es in Wirklichkeit grün ist: A-posteriori-Odds = (0,15/0,85) × (0,80/0,20) = 0,706. Die Odds sind der Quotient aus der Wahrscheinlichkeit, dass das Taxi blau ist, und der Wahrscheinlichkeit, dass das Taxi grün ist. Um die Wahrscheinlichkeit zu erhalten, dass das Taxi blau ist, berechnen wir: Wahrscheinlichkeit (blau) = 0,706 / 1,706 = 0,41. Die Wahrscheinlichkeit, dass das Taxi blau ist, beträgt 41 Prozent.

2
Amos Tversky und Daniel Kahneman, »Causal Schemas in Judgments Under Uncertainty«, in: *Progress in Social Psychology*, Morris Fishbein (Hg.), Hillsdale 1980, S. 49–72.

3
Richard E. Nisbett und Eugene Borgida, »Attribution and the Psychology of Prediction«, *Journal of Personality and Social Psychology* 32 (1975): S. 932–943.

4
John M. Darley und Bibb Latane, »Bystander Intervention in Emergencies: Diffusion of Responsibility«, *Journal of Personality and Social Psychology* 8 (1968): S. 377–383.

## 17. Regression zum Mittelwert

1
Michael Bulmer, *Francis Galton: Pioneer of Heredity and Biometry*, Baltimore 2003.

2
Wissenschaftler formen einen ursprünglichen Messwert in einen Standardwert um, indem sie den Mittelwert subtrahieren und das Ergebnis mit der Standardabweichung dividieren. Standardwerte haben einen Mittelwert von null und eine Standardabweichung von eins, können zwischen verschiedenen Variablen verglichen werden (insbesondere wenn die statistischen Verteilungen der ursprünglichen Werte ähnlich sind) und haben viele wünschenswerte mathematische Eigenschaften, die Galton herausarbeiten musste, um die Natur von Korrelation und Regression zu verstehen.

3
Dies gilt nicht in einem Umfeld, in dem einige Kinder unterernährt sind: Unterschiede in der Ernährung gewinnen an Bedeutung, der Anteil der gemeinsamen Faktoren nimmt ab und damit die Korrelation zwischen der Körpergröße der Eltern und der Körpergröße der Kinder (es sei denn, die Eltern unterernährter Kinder waren selbst aufgrund von Unterernährung in ihrer eigenen Kindheit minderwüchsig).

4
Die Korrelation wurde für eine sehr große Stichprobe aus der Bevölkerung der Vereinigten Staaten berechnet (der sog. *Gallup-Healthways Well-Being Index*).

5
Die Korrelation scheint eindrucksvoll zu sein, aber mit Erstaunen erfuhr ich vor vielen Jahren von dem Soziologen Christopher Jencks, dass, wenn alle das gleiche Bildungsniveau hätten, die Einkommensungleichheit (gemessen durch die Standardabweichung) nur um etwa 9 Prozent verringert würde. Die einschlägige Formel ist hier $\sqrt{(1-r^2)}$, wobei r die Korrelation ist.

6
Dies gilt, wenn beide Variablen in Standardwerten gemessen werden – das heißt, jeder Wert wird durch Subtraktion des Mittelwertes und Division des

Ergebnisses mit der Standardabweichung umgeformt.

7

Howard Wainer, »The Most Dangerous Equation«, *American Scientist* 95 (2007): S. 249–256.

## 18. Intuitive Vorhersagen bändigen

1

Der Beweis dafür, dass die Standardregression die optimale Lösung für das Vorhersageproblem ist, geht davon aus, dass Fehler mit der quadratischen Abweichung vom richtigen Wert gewichtet werden. Dies ist das allgemein anerkannte Kriterium der kleinsten Quadrate. Andere Verlustfunktionen führen zu anderen Lösungen.

# TEIL III

## 19. Die Illusion des Verstehens

1
Nassim Nicholas Taleb, *Der Schwarze Schwan: Die Macht höchst unwahrscheinlicher Ereignisse (The Black Swan: The Impact of the Highly Improbable)*, München 2008.
2
Vgl. Kapitel 7.
3
Michael Lewis, *Moneyball: The Art of Winning an Unfair Game*, New York 2003.
4
Seth Weintraub, »Excite Passed Up Buying Google for $ 750,000 in 1999«, *Fortune*, 29. September 2011.
5
Richard E. Nisbett und Timothy D. Wilson, »Telling More Than We Can Know: Verbal Reports on Mental Processes«, *Psychological Review* 84 (1977): S. 231–259.
6
Baruch Fischhoff und Ruth Beyth, »I Knew It Would Happen: Remembered Probabilities of Once Future Things«, *Organizational Behavior and Human Performance* 13 (1975): S. 1–16.
7
Jonathan Baron und John C. Hershey, »Outcome Bias in Decision Evaluation«, *Journal of Personality and Social Psychology* 54 (1988): S. 569–579.
8
Kim A. Kamin und Jeffrey Rachlinski, »Ex Post ≠ Ex Ante: Determining Liability in Hindsight«, *Law and Human Behavior* 19 (1995): S. 89–104. Jeffrey J. Rachlinski, »A Positive Psychological Theory of Judging in Hindsight«, *University of Chicago Law Review* 65 (1998): S. 571–625.
9
Jeffrey Goldberg, »Letter from Washington: Woodward vs. Tenet«, *New Yorker*, 21. Mai 2007, S. 35–38. Auch Tim Weiner, *CIA: die ganze Geschichte (Legacy of Ashes: The History of the CIA)*, Frankfurt a. M.

2008; »Espionage: Inventing the Dots«, *The Economist*, 3. November 2007, S. 100.

10

Philip E. Tetlock, »Accountability: The Neglected Social Context of Judgment and Choice«, *Research in Organizational Behavior* 7 (1985): S. 297–332.

11

Marianne Bertrand und Antoinette Schoar, »Managing with Style: The Effect of Managers on Firm Policies«, *Quarterly Journal of Economics* 118 (2003): S. 1169–1208. Nick Bloom und John Van Reenen, »Measuring and Explaining Management Practices Across Firms and Countries«, *Quarterly Journal of Economics* 122 (2007): S. 1351–1408.

12

Ich danke Herrn Professor James H. Steiger von der Vanderbilt University, der einen Algorithmus entwickelte, welcher diese Frage unter plausiblen Annahmen beantwortet. Steigers Analyse zeigt, dass Korrelationen von 0,20 und 0,40 mit Inversionsraten von 43 und 37 Prozent verbunden sind.

13

*Der Halo-Effekt* wurde sowohl von der *Financial Times* als auch vom *Wall Street Journal* als eines der besten Wirtschaftsbücher des Jahres gelobt: Philip Rosenzweig, *Der Halo-Effekt. Wie Manager sich täuschen lassen (The Halo Effect: … and the Eight Other Business Delusions That Deceive Managers)*, Offenbach 2008. Vgl. auch Paul Olk und Phil Rosenzweig, »The Halo Effect and the Challenge of Management Inquiry: A Dialog Between Phil Rosenzweig and Paul Olk«, *Journal of Management Inquiry* 19 (2010): S. 48–54.

14

James C. Collins und Jerry I. Porras, *Immer erfolgreich. Die Strategien der Topunternehmen (Built to Last: Successful Habits of Visionary Companies)*, Stuttgart, München 2003.

15

Sogar wenn Sie selbst der Vorstandschef wären, wären Ihre Prognosen nicht in einer eindrucksvollen Weise zutreffend; die umfassende Erforschung des Insiderhandels hat gezeigt, dass Manager beim Handel mit ihren eigenen Aktien den Markt schlagen, aber die Spanne ihrer Outperformance reicht kaum aus, um die Handelskosten zu decken. Vgl. H. Nejat Seyhun, »The Information Content of Aggregate Insider Trading«, *Journal of Business* 61

(1988): S. 1–24; Josef Lakonishok und Inmoo Lee, »Are Insider Trades Informative?«, *Review of Financial Studies* 14 (2001): S. 79–111; Zahid Iqbal und Shekar Shetty, »An Investigation of Causality Between Insider Transactions and Stock Returns«, *Quarterly Review of Economics and Finance* 42 (2002): S. 41–57.

16

Tom J. Peters und Robert H. Waterman, *Auf der Suche nach Spitzenleistungen (In Search of Excellence)*, München 2007. Rosenzweig, *Der Halo-Effekt*

17

Deniz Anginer, Kenneth L. Fisher und Meir Statman, »Stocks of Admired Companies and Despised Ones«, Arbeitspapier 2007.

18

Jason Zweig weist darauf hin, dass sich das mangelnde Verständnis des Phänomens der Regression negativ auf die Einstellung von Vorstandschefs auswirkt. Firmen, die ums Überleben kämpfen, greifen in der Regel auf externe Manager zurück und rekrutieren Vorstandschefs von Unternehmen, die in jüngster Vergangenheit sehr ertragsstark gewesen sind. Der neue Vorstandschef wird dann, zumindest vorübergehend, für die anschließende Verbesserung der Ertragslage seines neuen Unternehmens gelobt. (Inzwischen kämpft sein Nachfolger bei seinem früheren Unternehmen mit Problemen, was die Aufsichtsräte der neuen Firma zu der Überzeugung bringt, sie hätten zweifelsfrei »den richtigen Mann« eingestellt.) Jedes Mal, wenn ein Vorstandsvorsitzender »von Bord geht«, muss das neue Unternehmen seine Anteile (in Aktien und Bezugsrechten) an seiner alten Firma aufkaufen, was eine Basislinie für die zukünftige Vergütung festsetzt, die nichts mit der Leistung im neuen Unternehmen zu tun hat. Für »persönliche« Leistungen, hinter denen sich hauptsächlich Regressions- und Halo-Effekte verbergen, werden Erfolgsprämien in Höhe von vielen Millionen Dollar zuerkannt. (Persönliche Mitteilung, 29. Dezember 2009)

## 20. Die Illusion der Gültigkeit

1
Brad M. Barber und Terrance Odean, »Trading Is Hazardous to Your Wealth: The Common Stock Investment Performance of Individual Investors«, *Journal of Finance* 55 (2002): S. 773–806.

2
Brad M. Barber und Terrance Odean, »Boys Will Be Boys: Gender, Overconfidence, and Common Stock Investment«, *Quarterly Journal of Economics* 116 (2006): S. 261–292.

3
Dieser Dispositionseffekt wird in Kapitel 32 ausführlicher erörtert.

4
Brad M. Barber und Terrance Odean, »All That Glitters: The Effect of Attention and News on the Buying Behavior of Individual and Institutional Investors«, *Review of Financial Studies* 21 (2008): S. 785–818.

5
Eine Studie über den Aktienhandel in Taiwan gelangte zu der Schlussfolgerung, dass sich die Vermögensübertragung von Einzelpersonen auf Finanzinstitute auf sage und schreibe 2,2 Prozent des Bruttoinlandsprodukts beläuft: Brad M. Barber, Yi-Tsung Lee, Yu-Jane Liu und Terrance Odean, »Just How Much Do Individual Investors Lose by Trading?«, *Review of Financial Studies* 22 (2009): S. 609–632.

6
John C. Bogle, *Common Sense on Mutual Funds: New Imperatives for the Intelligent Investor,* New York 2000, S. 213.

7
Mark Grinblatt und Sheridan Titman, »The Persistence of Mutual Fund Performance«, *Journal of Finance* 42 (1992): S. 1977–1984. Edwin J. Elton et al., »The Persistence of Risk-Adjusted Mutual Fund Performance«, *Journal of Business* 52 (1997): S. 1–33. Edwin J. Elton et al., »Efficiency With Costly Information: A Reinterpretation of Evidence from Managed Portfolios«, *Review of Financial Studies* 6 (1993): S. 1–21.

8
Philip E. Tetlock, *Expert Political Judgment: How Good is It? How Can We Know?,* Princeton 2005, S. 233.

## 21. Intuitionen und Formeln

1
Paul Meehl, »Causes and Effects of My Disturbing Little Book«, *Journal of Personality Assessment* 50 (1986): S. 370–375.
2
Während der Auktionssaison 1990 bis 1991 zum Beispiel betrug der durchschnittliche Preis einer Kiste Château Latour Jahrgang 1960 in London 464 Dollar; eine Kiste des Jahrgangs 1961 (einer der besten, den es je gab) brachte durchschnittlich 5432 Dollar ein.
3
Paul J. Hoffman, Paul Slovic und Leonard G. Rorer, »An Analysis-of-Variance Model for the Assessment of Configural Cue Utilization in Clinical Judgment«, *Psychological Bulletin* 69 (1968): S. 338–339.
4
Paul R. Brown, »Independent Auditor Judgment in the Evaluation of Internal Audit Functions«, *Journal of Accounting Research* 21 (1983): S. 444–455.
5
James Shanteau, »Psychological Characteristics and Strategies of Expert Decision Makers«, *Acta Psychologica* 68 (1988): S. 203–215.
6
Danziger, Levav und Avnaim-Pesso, »Extraneous Factors in Judicial Decisions«.
7
Richard A. DeVaul et al., »Medical-School Performance of Initially Rejected Students«, *JAMA* 257 (1987): S. 47–51. Jason Dana und Robyn M. Dawes, »Belief in the Unstructured Interview: The Persistence of an Illusion«, Arbeitspapier, Department of Psychology, University of Pennsylvania, 2011. William M. Grove et al., »Clinical Versus Mechanical Prediction: A Meta-Analysis«, *Psychological Assessment* 12 (2000): S. 19–30.
8
Robyn M. Dawes, »The Robust Beauty of Improper Linear Models in Decision Making«, *American Psychologist* 34 (1979): S. 571–582.
9

Jason Dana und Robyn M. Dawes, »The Superiority of Simple Alternatives to Regression for Social Science Predictions«, *Journal of Educational and Behavioral Statistics* 29 (2004): S. 317–331.

10

Virginia Apgar, »A Proposal for a New Method of Evaluation of the Newborn Infant«, *Current Researches in Anesthesia and Analgesia* 32 (1953): S. 260–267. Mieczyslaw Finster und Margaret Wood, »The Apgar Score Has Survived the Test of Time«, *Anesthesiology* 102 (2005): S. 855–857.

11

Atul Gawande, *The Checklist Manifesto: How to Get Things Right*, New York 2009.

12

Paul Rozin, »The Meaning of ›Natural‹: Process More Important than Content«, *Psychological Science* 16 (2005): S. 652–658.

## 22. Die Intuition von Experten: Wann können wir ihr vertrauen?

1
Mellers, Hertwig und Kahneman, »Do Frequency Representations Eliminate Conjunction Effects?«.
2
Klein, *Natürliche Entscheidungsprozesse. Über die »Quellen der Macht«, die unsere Entscheidungen lenken* (*Sources of Power).*
3
Das J. Paul Getty Museum in Los Angeles beauftragt die führenden Experten der Welt auf dem Gebiet der altgriechischen Bildhauerkunst, einen Kouros – die Marmorstatue eines schreitenden Jungen –, die es zu kaufen erwägt, zu begutachten. Die Experten reagieren, einer nach dem anderen, mit dem, was man »intuitive Abneigung« nennen könnte – einem starken »Bauchgefühl«, dass der Kouros nicht 2500 Jahre alt ist, sondern eine moderne Fälschung. Keiner der Experten vermag auf Anhieb zu sagen, wieso er glaubt, dass es sich bei der Skulptur um eine Fälschung handelt. Der Hinweis eines italienischen Kunsthistorikers, dass irgendetwas – er könne nicht genau sagen, was – mit den Fingernägeln der Statue »nicht zu stimmen« scheine, war das Äußerste an einer rationalen Begründung, was ihnen unterkam. Ein berühmter amerikanischer Experte sagte, das Wort »neu« sei der erste Gedanke gewesen, der ihm bei der Betrachtung der Skulptur gekommen sei, und ein griechischer Experte behauptete doch glatt: »Jeder, der jemals eine ausgegrabene Skulptur gesehen hat, weiß, dass dieses Objekt nie in der Erde gelegen hat.« Die fehlende Einigkeit über die Gründe der gemeinsamen Schlussfolgerung ist bemerkenswert und eher verdächtig.
4
Simon war einer der überragenden Wissenschaftler des 20. Jahrhunderts. Er schrieb einen Klassiker über Entscheidungsfindung in Organisationen, als er noch in seinen Zwanzigern war, und neben seinen vielen anderen bedeutenden Leistungen war er einer der Begründer der Disziplin der Künstlichen Intelligenz, ein führender Kognitionsforscher, ein einflussreicher Erforscher des Prozesses wissenschaftlicher Erkenntnisgewinnung, ein Wegbereiter der Verhaltensökonomik und,

beinahe nebenher, auch noch Nobelpreisträger für Wirtschaftswissenschaften.

5

Simon, »What Is an Explanation of Behavior ?«. David G. Myers, *Intuition: Its Powers and Perils*, New Haven 2002, S. 56.

6

Seymour Epstein, »Demystifying Intuition: What It Is, What It Does, How It Does It«, *Psychological Inquiry* 21 (2010): S. 295–312.

7

Foer, *Moonwalk mit Einstein: Wie aus einem vergesslichen Mann ein Gedächtnis-Champion wurde* (*Moonwalking with Einstein).*

## 23. Die Außensicht

1
Diese Bezeichnungen werden oft missverstanden. Zahlreiche Autoren waren der Ansicht, die richtigen Termini seien »Insider-Sicht« und »Outsider-Sicht«, die nicht annähernd dem entsprachen, was uns vorschwebte.

2
Dan Lovallo und Daniel Kahneman, »Timid Choices and Bold Forecasts: A Cognitive Perspective on Risk Taking«, *Management Science* 39 (1993): S. 17–31. Daniel Kahneman und Dan Lovallo, »Delusions of Success: How Optimism Undermines Executives' Decisions«, *Harvard Business Review* 81 (2003): S. 56–63.

3
Richard E. Nisbett und Lee D. Ross, *Human Inference: Strategies and Shortcomings of Social Judgment*, Englewood Cliffs 1980.

4
Ein Beispiel für Zweifel an der sogenannten »evidenzbasierten« Medizin findet sich in Jerome Groopman, *How Doctors Think*, New York 2008, S. 6, dt. *Wie Ärzte ticken. Warum Mediziner so oder anders entscheiden*, München 2010

5
Daniel Kahneman und Amos Tversky, »Intuitive Prediction: Biases and Corrective Procedures«, *Management Science* 12 (1979): S. 313–327.

6
Rt. Hon. The Lord Fraser of Carmyllie, »The Holyrood Inquiry, Final Report«, 8. September 2004, www.holyroodinquiry.org/FINAL_report/report.htm

7
Brent Flyvbjerg, Mette K. Skamris Holm und Søren L. Buhl, »How (In)accurate Are Demand Forecasts in Public Works Projects ?«, *Journal of the American Planning Association* 71 (2005): S. 131–146.

8
»2002 Cost vs. Value Report«, *Remodeling*, 20. November 2002.

9

Brent Flyvbjerg, »From Nobel Prize to Project Management: Getting Risks Right«, *Project Management Journal* 37 (2006): S. 5–15.

10

Hal R. Arkes und Catherine Blumer, »The Psychology of Sunk Cost«, *Organizational Behavior and Human Decision Processes* 35 (1985): S. 124–140. Hal R. Arkes und Peter Ayton, »The Sunk Cost and Concorde Effects: Are Humans Less Rational Than Lower Animals?«, *Psychological Bulletin* 125 (1998): S. 591–600.

## 24. Die Maschine des Kapitalismus

1
Miriam A. Mosing et al., »Genetic and Environmental Influences on Optimism and Its Relationship to Mental and Self-Rated Health: A Study of Aging Twins«, *Behavior Genetics* 39 (2009): S. 597–604. David Snowdon, *Lieber alt und gesund. Dem Altern seinen Schrecken nehmen (Aging with Grace: What the Nun Study Teaches Us About Leading Longer, Healthier, and More Meaningful Lives)*, München 2001.

2
Elaine Fox, Anna Ridgewell und Chris Ashwin, »Looking on the Bright Side: Biased Attention and the Human Serotonin Transporter Gene«, *Proceedings of the Royal Society B* 276 (2009): S. 1747–1751.

3
Manju Puri und David T. Robinson, »Optimism and Economic Choice«, *Journal of Financial Economics* 86 (2007): S. 71–99.

4
Lowell W. Busenitz und Jay B. Barney, »Differences Between Entrepreneurs and Managers in Large Organizations: Biases and Heuristics in Strategic Decision-Making«, *Journal of Business Venturing* 12 (1997): S. 9–30.

5
Die vermutlich irrige Überzeugung, sie hätten eine Menge aus der Erfahrung gelernt, sorgt dafür, dass gescheiterte Unternehmer ihre Zuversicht nicht verlieren. Gavin Cassar und Justin Craig, »An Investigation of Hindsight Bias in Nascent Venture Activity«, *Journal of Business Venturing* 24 (2009): S. 149–164.

6
Keith M. Hmieleski und Robert A. Baron, »Entrepreneurs' Optimism and New Venture Performance: A Social Cognitive Perspective«, *Academy of Management Journal* 52 (2009): S. 473–488. Matthew L. A. Hayward, Dean A. Shepherd und Dale Griffin, »A Hubris Theory of Entrepreneurship«, *Management Science* 52 (2006): S. 160–172.

7
Arnold C. Cooper, Carolyn Y. Woo und William C. Dunkelberg, »Entrepreneurs' Perceived Chances for Success«, *Journal of Business*

*Venturing* 3 (1988): S. 97–108.

8

Thomas Åstebro und Samir Elhedhli, »The Effectiveness of Simple Decision Heuristics: Forecasting Commercial Success for Early-Stage Ventures«, *Management Science* 52 (2006): S. 395–409.

9

Thomas Åstebro, »The Return to Independent Invention: Evidence of Unrealistic Optimism, Risk Seeking or Skewness Loving?«, *Economic Journal* 113 (2003): S. 226–239.

10

Eleanor F. Williams und Thomas Gilovich, »Do People Really Believe They Are Above Average?«, *Journal of Experimental Social Psychology* 44 (2008): S. 1121–1128.

11

Richard Roll, »The Hubris Hypothesis of Corporate Takeovers«, *Journal of Business* 59 (1986): S. 197–216, Teil 1. Dieser bemerkenswerte frühe Aufsatz präsentierte eine verhaltenswissenschaftliche Analyse von Fusionen und Akquisitionen, die die Annahme der Rationalität preisgab, lange bevor solche Analysen populär wurden.

12

Ulrike Malmendier und Geoffrey Tate, »Who Makes Acquisitions? CEO Overconfidence and the Market's Reaction«, *Journal of Financial Economics* 89 (2008): S. 20–43.

13

Ulrike Malmendier und Geoffrey Tate, »Superstar CEOs«, *Quarterly Journal of Economics* 24 (2009), S. 1593–1638.

14

Paul D. Windschitl, Jason P. Rose, Michael T. Stalkfleet und Andrew R. Smith, »Are People Excessive or Judicious in Their Egocentrism? A Modeling Approach to Understanding Bias and Accuracy in People's Optimism«, *Journal of Personality and Social Psychology* 95 (2008): S. 252–273.

15

Eine Form der Konkurrenzvernachlässigung wurde auch zu jener Tageszeit beobachtet, zu der Verkäufer auf eBay ihre Auktionen beenden. Die leichte Frage lautet: Um wie viel Uhr ist die Gesamtzahl der Bieter am höchsten? Die Antwort: um 19 Uhr EST (Eastern Standard Time). Die Frage, die

Verkäufer beantworten sollten, ist schwieriger: Zu welcher Zeit sehen sich die meisten Bieter meine Auktion an, wenn man bedenkt, wie viele andere Verkäufer ihre Auktionen in Spitzenzeiten abschließen? Die Antwort: gegen Mittag, wenn die Anzahl der Bieter im Verhältnis zur Anzahl der Verkäufer groß ist. Die Verkäufer, die sich an die Konkurrenz erinnern und die Hauptauktionszeiten vermeiden, erzielen höhere Preise. Uri Simonsohn, »eBay's 's Crowded Evenings: Competition Neglect in Market Entry Decisions«, *Management Science* 56 (2010): S. 1060–1073.

16

Eta S. Berner und Mark L. Graber, »Overconfidence as a Cause of Diagnostic Error in Medicine«, *American Journal of Medicine* 121 (2008): S2 – S23.

17

Pat Croskerry und Geoff Norman, »Overconfidence in Clinical Decision Making«, *American Journal of Medicine* 121 (2008): S24 – S29.

18

Kahneman und Lovallo, »Timid Choices and Bold Forecasts«.

19

J. Edward Russo und Paul J. H. Schoemaker, »Managing Overconfidence«, *Sloan Management Review* 33 (1992): S. 7–17

# TEIL IV

## 25. Irrtümer

1
Clyde H. Coombs, Robyn M. Dawes und Amos Tversky, *Mathematische Psychologie. Eine Einführung (Mathematical Psychology: An Elementary Introduction)*, Weinheim, Basel 1975.

2
Diese Regel gilt näherungsweise für viele Dimensionen der Empfindung und Wahrnehmung. Sie wird nach ihrem Entdecker, dem deutschen Physiologen Ernst Heinrich Weber, »Webersches Gesetz« genannt. Fechner leitete aus dem Weberschen Gesetz die logarithmische psychophysische Funktion ab.

3
Bernoullis Intuition war richtig, und Ökonomen verwenden in vielen Kontexten noch immer den Logarithmus des Einkommens oder des Vermögens. Als zum Beispiel Angus Deaton die durchschnittliche Lebenszufriedenheit der Einwohner vieler Länder in Abhängigkeit vom Bruttoinlandsprodukt dieser Länder darstellte, verwendete er den Logarithmus des BIP als Einkommensmaß. Wie sich zeigte, ist der Zusammenhang ein sehr enger: Einwohner von Ländern mit hohem BIP sind mit der Qualität ihres Lebens viel zufriedener als Einwohner armer Länder, und eine Verdopplung des Einkommens geht in reichen und armen Ländern annähernd mit der gleichen Zunahme der Lebenszufriedenheit einher.

4
Nicholas Bernoulli, ein Cousin von Daniel Bernoulli, stellte eine Frage, die sich folgendermaßen paraphrasieren lässt: »Man lädt Sie zu einem Glücksspiel ein, bei dem Sie wiederholt eine Münze werfen. Bei ›Kopf‹ erhalten Sie zwei Dollar, und der Gewinn verdoppelt sich mit jedem nachfolgenden Münzwurf, bei dem sich ›Kopf‹ zeigt. Das Spiel endet, wenn die Münze zum ersten Mal ›Zahl‹ zeigt. Wie viel würden Sie bezahlen, um an dem Glücksspiel teilzunehmen?« Menschen glauben nicht, dass das Glücksspiel mehr wert ist als ein paar Dollar, obwohl sein Erwartungswert unendlich ist – weil der Gewinn, bei einem Erwartungswert von einem Dollar pro Münzwurf, ins Unendliche steigt. Der Nutzen der Gewinne dagegen erhöht sich viel langsamer, was erklärt, dass das Glücksspiel nicht

attraktiv ist. 5 Weitere Faktoren trugen dazu bei, dass sich Bernoullis Theorie so lange halten konnte. Ein Faktor besteht darin, dass Wahlen zwischen Glücksspielen wie selbstverständlich in Kategorien von Gewinnen, gemischten Gewinnen und Verlusten formuliert werden. Nicht viele Menschen dachten über Entscheidungen nach, bei denen alle Optionen schlecht sind, auch wenn wir keineswegs die Ersten waren, die risikofreudiges Verhalten beobachteten. Bernoullis Theorie wird außerdem dadurch begünstigt, dass es oftmals sehr vernünftig ist, die endgültigen Vermögenszustände zu betrachten und die Vergangenheit zu ignorieren. Ökonomen interessierten sich von jeher für rationale Entscheidungen, und Bernoullis Modell eignete sich gut für diesen Zweck.

## 26. Die Neue Erwartungstheorie

1
Stanley S. Stevens, »To Honor Fechner and Repeal His Law«, Science 133 (1961): S. 80–86. Stevens, *Psychophysics*.

2
Beim Schreiben dieses Satzes fiel mir ein, dass der Graph der Wertfunktion bereits als ein Emblem benutzt wurde. Jeder Nobelpreisträger erhält eine individuell gestaltete Urkunde mit einer personalisierten Zeichnung, die vermutlich von dem Nobelpreiskomitee ausgewählt wird. Meine Illustration war eine stilisierte Wiedergabe von Abbildung 10.

3
Die Verlustaversionsrate bewegt sich oft in einem Bereich zwischen 1,5 und 2,5: Nathan Novemsky und Daniel Kahneman, »The Boundaries of Loss Aversion«, Journal of Marketing Research 42 (2005): S. 119–128.

4
Peter Sokol-Hessner et al., »Thinking Like a Trader Selectively Reduces Individuals' Loss Aversion«, *PNAS* 106 (2009): S. 5035–5040.

5
Mehrere Jahre hintereinander hielt ich in dem finanzwirtschaftlichen Einführungskurs meines Kollegen Burton Malkiel eine Gastvorlesung. Jedes Jahr diskutierte ich auch die mangelnde Plausibilität von Bernoullis Theorie. Als ich Rabins Beweis zum ersten Mal erwähnte, bemerkte ich einen deutlichen Einstellungswandel bei meinem Kollegen. Er war jetzt bereit, die Schlussfolgerung viel ernster zu nehmen als in der Vergangenheit. Mathematische Argumente besitzen eine bestimmte Qualität, die zwingender ist als Appelle an den gesunden Menschenverstand. Ökonomen sind besonders empfänglich für diesen Vorteil.

6
Die intuitive Plausibilität des Beweises lässt sich an einem Beispiel verdeutlichen. Angenommen, das Vermögen einer Person ist W, und sie lehnt eine Lotterie mit gleichen Wahrscheinlichkeiten, 11 Dollar zu gewinnen oder 10 Dollar zu verlieren, ab. Wenn die Nutzenfunktion von Vermögen konkav (nach innen gewölbt) ist, bedeutet die Präferenz, dass der Wert von 1 Dollar über ein Intervall von 21 Dollar um über 9 Prozent

gesunken ist! Dies ist ein außerordentlich steiler Rückgang, und der Effekt verstärkt sich stetig in dem Maße, wie die Lotterien immer extremer werden.

7

Matthew Rabin, »Risk Aversion and Expected-Utility Theory: A Calibration Theorem«, *Econometrica* 68 (2000): S. 1281–1292. Matthew Rabin und Richard H. Thaler, »Anomalies: Risk Aversion«, Journal of Economic Perspectives 15 (2001): S. 219–232.

8

Mehrere Theoretiker haben verschiedene Version von Reue-Theorien vorgestellt, die auf der Annahme beruhen, dass Menschen vorwegnehmen können, wie sich die Optionen, die sich nicht verwirklichten, und/oder die Entscheidungen, die sie nicht trafen, auf ihre zukünftigen Erfahrungen auswirken werden: David E. Bell, »Regret in Decision Making Under Uncertainty«, Operations Research 30 (1982): S. 961–981. Graham Loomes und Robert Sugden, »Regret Theory: An Alternative to Rational Choice Under Uncertainty«, Economic Journal 92 (1982): S. 805–825. Barbara A. Mellers, »Choice and the Relative Pleasure of Consequences«, Psychological Bulletin 126 (2000): S. 910–924. Barbara A. Mellers, Alan Schwartz und Ilana Ritov, »Emotion-Based Choice«, Journal of Experimental Psychology: General 128 (1999): S. 332–345. Die Wahlen, die Entscheider zwischen riskanten Optionen treffen, hängen davon ab, ob sie erwarten, das Ergebnis des Glücksspiels zu kennen, das sie nicht wählten. Ilana Ritov, »Probability of Regret: Anticipation of Uncertainty Resolution in Choice«, Organizational Behavior and Human Decision Processes 66 (1966): S. 228–236.

## 27. Der Endowment-Effekt

1
Eine theoretische Analyse, die von einer Verlustaversion ausgeht, sagt einen deutlichen Knick der Indifferenzkurve im Referenzpunkt vorher: Amos Tversky und Daniel Kahneman, »Loss Aversion in Riskless Choice: A Reference- Dependent Model«, *Quarterly Journal of Economics* 106 (1991): S. 1039–1061. Jack Knetsch beobachtete diese Knicks in einer experimentellen Studie: »Preferences and Nonreversibility of Indifference Curves«, *Journal of Economic Behavior & Organization* 17 (1992): S. 131–139.

2
Alan B. Krueger und Andreas Mueller, »Job Search and Job Finding in a Period of Mass Unemployment: Evidence from High-Frequency Longitudinal Data«, Arbeitspapier, Princeton University Industrial Relations Section, Januar 2011.

3
Streng genommen erlaubt die Theorie aufgrund des sogenannten »Einkommenseffekts«, dass der Kaufpreis geringfügig niedriger ist als der Verkaufspreis: Der Käufer und der Verkäufer sind nicht gleich vermögend, weil der Verkäufer eine Flasche mehr hat. Aber der Effekt ist in diesem Fall vernachlässigbar, weil 50 Dollar ein winziger Bruchteil des Vermögens des Professors sind. Die Theorie würde vorhersagen, dass dieser Einkommenseffekt seine Zahlungsbereitschaft nicht einmal um einen Penny verändern würde.

4
Der Ökonom Alan Krueger berichtete über eine Studie, die er durchführte, als er mit seinem Vater zum Super-Bowl-Finale der US-Football-Liga ging: »Wir fragten Fans, die bei einer Lotterie das Recht gewonnen hatten, ein paar Eintrittskarten für jeweils 325 oder 400 Dollar zu kaufen, ob sie bereit gewesen wären, 3000 Dollar für eine Eintrittskarte zu zahlen, wenn sie bei der Lotterie verloren hätten, und ob sie ihre Tickets verkauft hätten, wenn ihnen jemand 3000 Dollar pro Stück angeboten hätte. 94 Prozent sagten, sie hätten für 3000 Dollar nicht gekauft, und 92 Prozent äußerten, sie hätten zu diesem Preis nicht verkauft.« Er folgert daraus, dass »Rationalität beim Super Bowl Mangelware ist«. Alan B. Krueger, »Supply and Demand: An

Economist Goes to the Super Bowl«, *Milken Institute Review: A Journal of Economic Policy* 3 (2001): S. 22–29.

5

Streng genommen bezieht sich die Verlustaversion auf die antizipierte Lust und Unlust, die Entscheidungen bestimmen. Diese Antizipationen können in manchen Fällen falsch sein. Deborah A. Kermer et al., »Loss Aversion Is an Affective Forecasting Error«, *Psychological Science* 17 (2006): S. 649–653.

6

Novemsky und Kahneman, »The Boundaries of Loss Aversion«.

7

Stellen Sie sich vor, dass alle Teilnehmer gemäß dem ihnen zugeschriebenen Rückzahlungswert in einer Reihe angeordnet sind. Geben Sie jetzt an die Hälfte der Individuen in der Reihe nach dem Zufallsprinzip Gutscheine aus. Die Hälfte der Personen an der Spitze der Reihe hat keinen Gutschein, und die Hälfe der Personen am Ende der Reihe hat einen Gutschein. Diese Personen (die Hälfte der Gesamtzahl) werden sich wahrscheinlich bewegen, indem sie ihre Plätze tauschen, sodass am Schluss jeder in der ersten Hälfte der Reihe und niemand hinter ihnen einen Gutschein besitzt.

8

Brian Knutson et al., »Neural Antecedents of the Endowment Effect«, *Neuron* 58 (2008): S. 814–822. Brian Knutson und Stephanie M. Greer, »Anticipatory Affect: Neural Correlates and Consequences for Choice«, *Philosophical Transactions of the Royal Society B* 363 (2008): S. 3771–3786.

9

Eine Übersicht über den Risikopreis, basierend auf »internationalen Daten aus 16 verschiedenen Ländern über einen Zeitraum von mehr als 100 Jahren«, ergab einen Schätzwert von 2,3, »in verblüffender Übereinstimmung mit Schätzwerten, die mit der ganz anderen Methode von Laborexperimenten über individuelle Entscheidungsfindung gewonnen wurden«: Moshe Levy, »Loss Aversion and the Price of Risk«, *Quantitative Finance* 10 (2010): S. 1009–1022.

10

Miles O. Bidwel, Bruce X. Wang und J. Douglas Zona, »An Analysis of Asymmetric Demand Response to Price Changes: The Case of Local

Telephone Calls«, *Journal of Regulatory Economics* 8 (1995): S. 285–298. Bruce G. S. Hardie, Eric J. Johnson und Peter S. Fader, »Modeling Loss Aversion and Reference Dependence Effects on Brand Choice«, *Marketing Science* 12 (1993): S. 378–394.

11

Colin Camerer, »Three Cheers – Psychological, Theoretical, Empirical – for Loss Aversion«, *Journal of Marketing Research* 42 (2005): S. 129–133. Colin F. Camerer, »Prospect Theory in the Wild: Evidence from the Field«, in *Choices, Values, and Frames*, hg. von Daniel Kahneman und Amos Tversky, New York 2000, S. 288–300.

12

David Genesove und Christopher Mayer, »Loss Aversion and Seller Behavior: Evidence from the Housing Market«, *Quarterly Journal of Economics* 116 (2001): S. 1233–1260.

13

John A. List, »Does Market Experience Eliminate Market Anomalies?«, *Quarterly Journal of Economics* 118 (2003): S. 47–71.

14

Jack L. Knetsch, »The Endowment Effect and Evidence of Nonreversible Indifference Curves«, *American Economic Review* 79 (1989): S. 1277–1284.

15

Charles R. Plott und Kathryn Zeiler, »The Willingness to Pay – Willingness to Accept Gap, the ›Endowment Effect‹, Subject Misconceptions, and Experimental Procedures for Eliciting Valuations«, *American Economic Review* 95 (2005): S. 530–545. Charles Plott, ein führender Experimentalökonom, stand dem Endowment-Effekt sehr skeptisch gegenüber und versuchte zu zeigen, dass er kein »fundamentaler Aspekt der menschlichen Präferenzen« sei, sondern vielmehr das Ergebnis einer mangelhaften Technik. Plott und Zeiler glauben, dass Probanden, die den *Endowment-Effekt* zeigen, falsche Vorstellungen über die wahren Werte ihrer Güter haben, und sie modifizierten die Versuchsanordnungen der ursprünglichen Experimente, um diese Fehlannahmen zu beseitigen. Sie erarbeiteten ein ausgeklügeltes Trainingsverfahren, bei dem die Teilnehmer in die Rollen von Käufern und Verkäufern schlüpften und ihnen ausführlich beigebracht wurde, wie man die wahren Werte der Güter ermittelt. Wie zu erwarten, verschwand der Endowment-Effekt. Plott und Zeiler sehen in

ihrer Methode eine deutliche Verbesserung der experimentellen Technik. Psychologen würden dieser Methode erhebliche Mängel bescheinigen, weil den Probanden mitgeteilt wird, was die Experimentatoren als das angemessene Verhalten erachten, das »zufälligerweise« mit der Theorie der Experimentatoren übereinstimmt. Plott und Zeilers bevorzugte Version von Knetschs Tauschexperiment ist ähnlich tendenziös: Sie erlaubt dem Eigentümer des Gutes nicht, dieses physisch zu besitzen, was von entscheidender Bedeutung für den Effekt ist. Vgl. Charles R. Plott und Kathryn Zeiler, »Exchange Asymmetries Incorrectly Interpreted as Evidence of Endowment Effect Theory and Prospect Theory?«, *American Economic Review* 97 (2007): S. 1449–1466. Möglicherweise handelt es sich um eine Sackgasse, wo jede Seite die von der anderen Seite benötigten Methoden verwirft.

16

Bei ihren Studien über Entscheidungsfindung unter Armut haben Eldar Shafir, Sendhil Mullainathan und ihre Kollegen andere Fälle beobachtet, bei denen Armut ein ökonomisches Verhalten auslöst, das in mancher Hinsicht realistischer und rationaler ist als das von wohlhabenderen Menschen. Die Armen reagieren eher auf reale Ergebnisse als auf deren Beschreibung. Marianne Bertrand, Sendhil Mullainathan und Eldar Shafir, »Behavioral Economics and Marketing in Aid of Decision Making Among the Poor«, *Journal of Public Policy & Marketing* 25 (2006): S. 8–23.

17

Die Schlussfolgerung, dass für Anschaffungen ausgegebenes Geld nicht als ein Verlust erlebt wird, trifft mit höherer Wahrscheinlichkeit auf Menschen zu, die relativ wohlhabend sind. Entscheidend ist vermutlich die Frage, ob man sich beim Kauf eines Gutes der Tatsache bewusst ist, dass man sich auch weiterhin ein anderes Gut leisten kann. Novemsky und Kahneman, »The Boundaries of Loss Aversion«. Ian Bateman et al., »Testing Competing Models of Loss Aversion: An Adversarial Collaboration«, *Journal of Public Economics* 89 (2005): S. 1561–1580.

## 28. Negative Ereignisse

1
Paul J. Whalen et al., »Human Amygdala Responsivity to Masked Fearful Eye Whites«, *Science* 306 (2004): S. 2061. Personen mit fokalen Läsionen der Amygdala zeigten bei ihren risikobehafteten Entscheidungen nur eine geringe oder gar keine Verlustaversion: Benedetto De Martino, Colin F. Camerer und Ralph Adolphs, »Amygdala Damage Eliminates Monetary Loss Aversion«, *PNAS* 107 (2010): S. 3788–3792.

2
Joseph LeDoux, *Das Netz der Gefühle. Wie Emotionen entstehen. (The Emotional Brain: The Mysterious Underpinnings of Emotional Life)*, München 1998.

3
Elaine Fox et al., »Facial Expressions of Emotion: Are Angry Faces Detected More Efficiently?«, *Cognition & Emotion* 14 (2000): S. 61–92.

4
Christine Hansen und Ranald Hansen, »Finding the Face in the Crowd: An Anger Superiority Effect«, *Journal of Personality and Social Psychology* 54 (1988): S. 917–924.

5
Jos J. A. Van Berkum et al., »Right or Wrong? The Brain's Fast Response to Morally Objectionable Statements«, *Psychological Science* 20 (2009): S. 1092–1099.

6
Paul Rozin und Edward B. Royzman, »Negativity Bias, Negativity Dominance, and Contagion«, *Personality and Social Psychology Review* 5 (2001): S. 296–320.

7
Roy F. Baumeister, Ellen Bratslavsky, Catrin Finkenauer und Kathleen D. Vohs, »Bad Is Stronger Than Good«, *Review of General Psychology* 5 (2001): 323.

8
Michel Cabanac, »Pleasure: The Common Currency«, *Journal of Theoretical Biology* 155 (1992): S. 173–200.

9

Chip Heath, Richard P. Larrick und George Wu, »Goals as Reference Points«, *Cognitive Psychology* 38 (1999): S. 79–109.

10

Colin Camerer, Linda Babcock, George Loewenstein und Richard Thaler, »Labor Supply of New York City Cabdrivers: One Day at a Time«, *Quarterly Journal of Economics* 112 (1997): S. 407–441. Die Schlussfolgerungen aus diesen Forschungen wurden infrage gestellt: Henry S. Farber, »Is Tomorrow Another Day? The Labor Supply of New York Cab Drivers«, NBER Working Paper 9706, 2003. Eine Reihe von Studien über Fahrradkuriere in Zürich liefert starke Belege für die Wirksamkeit von Zielen, die sich mit den Ergebnissen der ursprünglichen Studie über Taxifahrer decken: Ernst Fehr und Lorenz Goette, »Do Workers Work More if Wages Are High? Evidence from a Randomized Field Experiment«, *American Economic Review* 97 (2007): S. 298–317.

11

Daniel Kahneman, »Reference Points, Anchors, Norms, and Mixed Feelings«, *Organizational Behavior and Human Decision Processes* 51 (1992): S. 296–312.

12

John Alcock, *Das Verhalten der Tiere aus evolutionsbiologischer Sicht (Animal Behavior: An Evolutionary Approach)*, Stuttgart 1996, S. 278–284, zitiert nach der amerik. Ausgabe (Sunderland 2009) von Eyal Zamir, »Law and Psychology: The Crucial Role of Reference *Points* and Loss Aversion«, Arbeitspapier, Hebrew University, 2011.

13

Daniel Kahneman, Jack L. Knetsch und Richard H. Thaler, »Fairness as a Constraint on Profit Seeking: Entitlements in the Market«, *The American Economic Review* 76 (1986): S. 728–741.

14

Ernst Fehr, Lorenz Goette und Christian Zehnder, »A Behavioral Account of the Labor Market: The Role of Fairness Concerns«, *Annual Review of Economics* 1 (2009): S. 355–384. Eric T. Anderson und Duncan I. Simester, »Price Stickiness and Customer Antagonism«, *Quarterly Journal of Economics* 125 (2010): S. 729–765.

15

Dominique de Quervain et al., »The Neural Basis of Altruistic Punishment«, *Science* 305 (2004): S. 1254–1258.

16
David Cohen und Jack L. Knetsch, »Judicial Choice and Disparities Between Measures of Economic Values«, *Osgoode Hall Law Review* 30 (1992): S. 737–770. Russell Korobkin, »The Endowment Effect and Legal Analysis«, *Northwestern University Law Review* 97 (2003): S. 1227–1293.
17
Zamir, »Law and Psychology«.

## 29. Das viergeteilte Muster

1
Einschließlich der Bekanntschaft mit einem sogenannten »Dutch Book«, bei dem es sich um eine Serie von Glücksspielen handelt, die man aufgrund seiner fehlerhaften Präferenzen akzeptiert und die zwangsläufig in einem Verlust enden.

2
Leser, die mit dem Allais-Paradoxon vertraut sind, werden erkennen, dass es sich hier um eine neue Version handelt. Sie ist viel einfacher als das ursprüngliche Paradoxon und verstößt zugleich viel stärker gegen die Axiome rationaler Wahl. Beim ersten Problem wird die linke Option bevorzugt. Das zweite Problem erhält man dadurch, dass man die linke Option um eine wertvollere Aussicht ergänzt als die rechte; trotzdem wird die rechte Option vorgezogen.

3
Wie der bekannte Wirtschaftswissenschaftler Kenneth Arrow die Konferenz unlängst beschrieb, schenkten deren Teilnehmer dem, was er »Allais' kleines Experiment« nannte, kaum Beachtung. Persönliche Mitteilung, 16. März 2011.

4
Die Tabelle zeigt Entscheidungsgewichte für Gewinne. Schätzungen für Verluste waren ganz ähnlich.

5
Ming Hsu, Ian Krajbich, Chen Zhao und Colin F. Camerer, »Neural Response to Reward Anticipation under Risk Is Nonlinear in Probabilities«, *Journal of Neuroscience* 29 (2009): S. 2231–2237.

6
W. Kip Viscusi, Wesley A. Magat und Joel Huber, »An Investigation of the Rationality of Consumer Valuations of Multiple Health Risks«, *RAND Journal of Economics* 18 (1987): S. 465–479.

7
In einem rationalen Modell mit abnehmendem Grenznutzen sollten Menschen, um die Häufigkeit von Unfällen von 15 auf 5 Einheiten zu verringern, mindestens zwei Drittel des Betrages zahlen, den sie für die

Beseitigung des Risikos zu zahlen bereit wären. Beobachtete Präferenzen verstießen gegen diese Vorhersage.

8

C. Arthur Williams, »Attitudes Toward Speculative Risks as an Indicator of Attitudes Toward Pure Risks«, *Journal of Risk and Insurance* 33 (1966): S. 577–586. Howard Raiffa, *Einführung in die Entscheidungstheorie (Decision Analysis: Introductory Lectures on Choices under Uncertainty)*, München, Wien 1973.

9

Chris Guthrie, »Prospect Theory, Risk Preference, and the Law«, *Northwestern University Law Review* 97 (2003): S. 1115–1163. Jeffrey J. Rachlinski, »Gains, Losses and the Psychology of Litigation«, *Southern California Law Review* 70 (1996): S. 113–185. Samuel R. Gross and Kent D. Syverud, »Getting to No: A Study of Settlement Negotiations and the Selection of Cases for Trial«, *Michigan Law Review* 90 (1991): S. 319–393. 10 Chris Guthrie, »Framing Frivolous Litigation: A Psychological Theory«, *University of Chicago Law Review* 67 (2000): S. 163–216.

## 30. Seltene Ereignisse

1
George F. Loewenstein, Elke U. Weber, Christopher K. Hsee und Ned Welch, »Risk as Feelings«, *Psychological Bulletin* 127 (2001): S. 267–286.
2
Ebd. Cass R. Sunstein, »Probability Neglect: Emotions, Worst Cases, and Law«, *Yale Law Journal* 112 (2002): S. 61–107. Siehe Anmerkungen zu Kapitel 13: Damasio, *Descartes' Irrtum*. Slovic, Finucane, Peters und MacGregor, »The Affect Heuristic«.
3
Craig R. Fox, »Strength of Evidence, Judged Probability, and Choice Under Uncertainty«, *Cognitive Psychology* 38 (1999): S. 167–189.
4
Urteile über die Wahrscheinlichkeiten eines Ereignisses und seines Komplements addieren sich nicht immer zu 100 Prozent. Wenn Menschen nach einem Thema gefragt werden, über das sie sehr wenig wissen (»Wie hoch schätzen Sie die Wahrscheinlichkeit ein, dass die Temperatur in Bangkok morgen Mittag über 38 Grad Celsius betragen wird?«), addieren sich die beurteilten Wahrscheinlichkeiten des Ereignisses und seines Komplements auf weniger als 100 Prozent.
5
In der Kumulativen Neuen Erwartungstheorie wird anders als in der ursprünglichen Version der Neuen Erwartungstheorie, die ich beschreibe, nicht davon ausgegangen, dass die Entscheidungsgewichte für Gewinne und Verluste gleich sind.
6
Die Frage über die beiden Urnen wurde erfunden von Dale T. Miller, William Turnbull und Cathy McFarland, »When a Coincidence Is Suspicious: The Role of Mental Simulation«, *Journal of Personality and Social Psychology* 57 (1989): S. 581–589. Seymour Epstein und seine Kollegen sprachen sich für eine Interpretation auf der Basis von zwei Systemen aus: Lee A. Kirkpatrick und Seymour Epstein, »Cognitive-Experiential Self-Theory and Subjective Probability: Evidence for Two Conceptual Systems«, *Journal of Personality and Social Psychology* 63 (1992): S. 534–544.

7
Kimihiko Yamagishi, »When a 12.86% Mortality Is More Dangerous Than 24.14%: Implications for Risk Communication«, *Applied Cognitive Psychology* 11 (1997): S. 495–506.
8
Slovic, Monahan und MacGregor, »Violence Risk Assessment and Risk Communication«.
9
Jonathan J. Koehler, »When Are People Persuaded by DNA Match Statistics?«, *Law and Human Behavior* 25 (2001): S. 493–513.
10
Ralph Hertwig, Greg Barron, Elke U. Weber und Ido Erev, »Decisions from Experience and the Effect of Rare Events in Risky Choice«, *Psychological Science* 15 (2004): S. 534–539. Ralph Hertwig und Ido Erev, »The Description-Experience Gap in Risky Choice«, *Trends in Cognitive Sciences* 13 (2009): S. 517–523.
11
Liat Hadar und Craig R. Fox, »Information Asymmetry in Decision from Description Versus Decision from Experience«, *Judgment and Decision Making* 4 (2009): S. 317–325.
12
Hertwig und Erev, »The Description-Experience Gap«.

## 31. Risikostrategien

1
Die Berechnung ist einfach. Jede der beiden Kombinationen besteht aus einer sicheren Option und einer Lotterie. Fügt man zu beiden Optionen der Lotterie die sichere Option hinzu, erhält man AD und BC.
2
Thomas Langer und Martin Weber, »Myopic Prospect Theory vs. Myopic Loss Aversion: How General Is the Phenomenon?«, *Journal of Economic Behavior & Organization* 56 (2005): S. 25–38.

## 32. Buch führen

1
Die Intuition wurde in einem Feldexperiment bestätigt, bei dem zufällig ausgewählte Studenten, die Jahresabos für das Hochschultheater kauften, ihre Eintrittskarten zu einem stark reduzierten Preis bekamen. Bei der anschließenden Erfassung der Besucher zeigte sich, dass Studenten, die für ihre Eintrittskarten den vollen Preis gezahlt hatten, mit höherer Wahrscheinlichkeit die Aufführungen besuchten, insbesondere in der ersten Hälfte der Spielzeit. Das Versäumen einer Vorstellung, für die man bezahlt hat, geht mit der unangenehmen Erfahrung einher, ein Konto mit Verlust schließen zu müssen. Arkes und Blumer, »The Psychology of Sunk Costs«.

2
Hersh Shefrin und Meir Statman, »The Disposition to Sell Winners Too Early and Ride Losers Too Long: Theory and Evidence«, *Journal of Finance* 40 (1985): S. 777–790. Terrance Odean, »Are Investors Reluctant to Realize Their Losses?«, *Journal of Finance* 53 (1998): S. 1775–1798.

3
Ravi Dhar und Ning Zhu, »Up Close and Personal: Investor Sophistication and the Disposition Effect«, *Management Science* 52 (2006): S. 726–740.

4
Darrin R. Lehman, Richard O. Lempert und Richard E. Nisbett, »The Effects of Graduate Training on Reasoning: Formal Discipline and Thinking about Everyday-Life Events«, *American Psychologist* 43 (1988): S. 431–442.

5
Marcel Zeelenberg und Rik Pieters, »A Theory of Regret Regulation 1.0«, *Journal of Consumer Psychology* 17 (2007): S. 3–18.

6
Kahneman und Miller, »Norm Theory«.

7
Die Anhalter-Frage wurde angeregt von einem berühmten Beispiel, das die Rechtsphilosophen Hart und Honoré diskutierten: »Eine Frau, die mit einem Mann verheiratet ist, der an einem Magengeschwür leidet, wird vielleicht den Verzehr von Pastinaken als Ursache seiner Verdauungsstörung ausmachen. Der Arzt wird vielleicht das

Magengeschwür als die Ursache und die Speise als einen bloßen Auslöser identifizieren.« Ungewöhnliche Ereignisse verlangen nach kausalen Erklärungen und rufen außerdem kontrafaktische Gedanken hervor, und die beiden sind eng miteinander verbunden. Dasselbe Ereignis kann entweder mit einer persönlichen Norm oder der Norm anderer Menschen verglichen werden, was zu verschiedenen kontrafaktischen Gedanken, verschiedenen Kausalattributionen und verschiedenen Emotionen (Reue oder Schuldgefühl) führen kann: Herbert L. A. Hart und Tony Honoré, *Causation in the Law*, New York 1985, S. 33.

8

Daniel Kahneman und Amos Tversky, »The Simulation Heuristic«, in *Judgment Under Uncertainty*: *Heuristics and Biases*, hg. von Daniel Kahneman, Paul Slovic und Amos Tversky, New York 1982, S. 160–173.

9

Janet Landman, »Regret and Elation Following Action and Inaction: Affective Responses to Positive Versus Negative Outcomes«, *Personality and Social Psychology Bulletin* 13 (1987): S. 524–536. Faith Gleicher et al., »The Role of Counterfactual Thinking in Judgments of Affect«, *Personality and Social Psychology Bulletin* 16 (1990): S. 284–295.

10

Dale T. Miller und Brian R. Taylor, »Counterfactual Thought, Regret, and Superstition: How to Avoid Kicking Yourself«, in *What Might Have Been: The Social Psychology of Counterfactual Thinking*, hg. von Neal J. Roese und James M. Olson, Hillsdale 1995, S. 305–331.

11

Marcel Zeelenberg, Kees van den Bos, Eric van Dijk und Rik Pieters, »The Inaction Effect in the Psychology of Regret«, *Journal of Personality and Social Psychology* 82 (2002): S. 314–3 27.

12

Itamar Simonson, »The Influence of Anticipating Regret and Responsibility on Purchase Decisions«, *Journal of Consumer Research* 19 (1992): S. 105–118.

13

Lilian Ng und Qinghai Wang, »Institutional Trading and the Turn-of-the-Year Effect«, *Journal of Financial Economics* 74 (2004): S. 343–366.

14

Tversky und Kahneman, »Loss Aversion in Riskless Choice«. Eric J. Johnson, Simon Gächter und Andreas Herrmann, »Exploring the Nature of Loss Aversion«, *Centre for Decision Research and Experimental Economics, University of Nottingham, Discussion Paper Series*, 2006. Edward J. McCaffery, Daniel Kahneman und Matthew L. Spitzer, »Framing the Jury: Cognitive Perspectives on Pain and Suffering«, *Virginia Law Review* 81 (1995): S. 1341–1420.

15

Richard H. Thaler, »Toward a Positive Theory of Consumer Choice«, *Journal of Economic Behavior and Organization* 39 (1980): S. 36–90.

16

Philip E. Tetlock et al., »The Psychology of the Unthinkable: Taboo Trade-Offs, Forbidden Base Rates, and Heretical Counterfactuals«, *Journal of Personality and Social Psychology* 78 (2000): S. 853–870.

17

Cass R. Sunstein, *The Laws of Fear: Beyond the Precautionary Principle*, New York 2005.

18

Daniel T. Gilbert et al., »Looking Forward to Looking Backward: The Misprediction of Regret«, *Psychological Science* 15 (2004): S. 346–350.

## 33. Umkehrungen

1
Dale T. Miller und Cathy McFarland, »Counterfactual Thinking and Victim Compensation: A Test of Norm Theory«, *Personality and Social Psychology Bulletin* 12 (1986): S. 513–519.

2
Den ersten Schritt zur gegenwärtigen Interpretation machten Max H. Bazerman, George F. Loewenstein und Sally B. White, »Reversals of Preference in Allocation Decisions: Judging Alternatives Versus Judging Among Alternatives«, *Administrative Science Quarterly* 37 (1992): S. 220–240. Christopher Hsee führte die Konzepte der gemeinsamen und getrennten Beurteilung ein und formulierte die wichtige Bewertbarkeitshypothese, die Umkehrungen mit der Annahme erklärt, dass manche Merkmale nur in der gemeinsamen Beurteilung bewertbar werden: »Attribute Evaluability: Its Implications for Joint-Separate Evaluation Reversals and Beyond«, in Kahneman und Tversky, *Choices, Values, and Frames*.

3
Sarah Lichtenstein und Paul Slovic, »Reversals of Preference Between Bids and Choices in Gambling Decisions«, *Journal of Experimental Psychology* 89 (1971): S. 46–55. Unabhängig davon gelangte Harold R. Lindman zu einem ähnlichen Ergebnis, »Inconsistent Preferences Among Gambles«, *Journal of Experimental Psychology* 89 (1971): S. 390–397.

4
Für eine Abschrift des berühmten Interviews vgl. Sarah Lichtenstein und Paul Slovic (Hg.), *The Construction of Preference*, New York 2006.

5
David M. Grether und Charles R. Plott, »Economic Theory of Choice and the Preference Reversals Phenomenon«, *American Economic Review* 69 (1979): S. 623–628.

6
Lichtenstein und Slovic, *The Construction of Preference*, S. 96.

7
Kuhn behauptete bekanntlich das Gleiche mit Bezug auf die Naturwissenschaften: Thomas S. Kuhn, »The Function of Measurement in

Modern Physical Science«, *Isis* 52 (1961): S. 161–193.

8

Es gibt Hinweise darauf, dass Fragen nach der emotionalen Anziehungskraft von Tierarten und die Bereitschaft, für ihren Schutz zu spenden, die gleichen Rangfolgen ergeben: Daniel Kahneman und Ilana Ritov, »Determinants of Stated Willingness to Pay for Public Goods: A Study in the Headline Method«, *Journal of Risk and Uncertainty* 9 (1994): S. 5–38.

9

Hsee, »Attribute Evaluability«.

10

Cass R. Sunstein, Daniel Kahneman, David Schkade und Ilana Ritov, »Predictably Incoherent Judgments«, *Stanford Law Review* 54 (2002): S. 1190.

## 34. Frames und Wirklichkeit

1
Amos Tversky und Daniel Kahneman, »The Framing of Decisions and the Psychology of Choice«, *Science* 211 (1981): S. 453–458.
2
Thaler, »Toward a Positive Theory of Consumer Choice«.
3
Barbara McNeil, Stephen G. Pauker, Harold C. Sox Jr. und Amos Tversky, »On the Elicitation of Preferences for Alternative Therapies«, *New England Journal of Medicine* 306 (1982): S. 1259–1262.
4
Einige Leute kritisierten, die Bezeichnung »asiatisch« sei unnötig und abwertend. Heute würden wir dieses Adjektiv vermutlich nicht mehr verwenden, aber das Beispiel wurde in den 1970er-Jahren geschrieben, als man weniger Sensibilität für Gruppenbezeichnungen besaß als heute. Das Wort wurde hinzugefügt, um das Beispiel konkreter zu machen, denn die Befragten wurden dadurch an die Epidemie der Asiatischen Grippe von 1957 erinnert.
5
Thomas Schelling, *Choice and Consequence*, Cambridge, 1985.
6
Richard P. Larrick und Jack B. Soll, »The MPG Illusion«, *Science* 320 (2008): S. 1593–1594.
7
Eric J. Johnson und Daniel Goldstein, »Do Defaults Save Lives?«, *Science* 302 (2003): S. 1338–1339.

# TEILV

## 35. Zwei Selbste

1
Vgl. O. Höffe (Hg.), *Einführung in die utilitaristische Ethik*, Tübingen 1992, S. 35.
2
Irving Fisher, »Is ›Utility‹ the Most Suitable Term for the Concept It Is Used to Denote?«, *American Economic Review* 8 (1918): S. 335.
3
Francis Edgeworth, *Mathematical Psychics*, New York 1881.
4
Daniel Kahneman, Peter P. Wakker und Rakesh Sarin, »Back to Bentham? Explorations of Experienced Utility«, *Quarterly Journal of Economics* 112 (1997): S. 375–405. Daniel Kahneman, »Experienced Utility and Objective Happiness: A Moment-Based Approach« und »Evaluation by Moments: Past and Future«, in Kahneman und Tversky, *Choices, Values, and Frames*, S. 673–692 und 693–708.
5
Donald A. Redelmeier und Daniel Kahneman, »Patients' Memories of Painful Medical Treatments: Real-time and Retrospective Evaluations of Two Minimally Invasive Procedures«, *Pain* 66 (1996): S. 3–8.
6
Daniel Kahneman, Barbara L. Frederickson, Charles A. Schreiber und Donald A. Redelmeier, »When More Pain Is Preferred to Less: Adding a Better End«, *Psychological Science* 4 (1993): S. 401–405.
7
Orval H. Mowrer und L. N. Solomon, »Contiguity vs. Drive-Reduction in Conditioned Fear: The Proximity and Abruptness of Drive Reduction«, *American Journal of Psychology* 67 (1954): S. 15–25.
8
Peter Shizgal, »On the Neural Computation of Utility: Implications from Studies of Brain Stimulation Reward«, in *Well-Being: The Foundations of Hedonic Psychology*, hg. von Daniel Kahneman, Ed Diener und Norbert Schwarz, New York 1999, S. 500–524.

## 36. Das Leben als eine Geschichte

1
Paul Rozin und Jennifer Stellar, »Posthumous Events Affect Rated Quality and Happiness of Lives«, *Judgment and Decision Making* 4 (2009): S. 273–279.

2
Ed Diener, Derrick Wirtz und Shigehiro Oishi, »End Effects of Rated Life Quality: The James Dean Effect«, *Psychological Science* 12 (2001): S. 124–128. Bei derselben Serie von Experimenten wurde auch die Höchststand-Ende-Regel in einem unglücklichen Leben überprüft, mit ähnlichen Ergebnissen: Jen wurde nicht als doppelt so unglücklich beurteilt, wenn ihr elendes Leben sechzig statt dreißig Jahre dauerte, aber die Glücksbilanz ihres Lebens fiel erheblich besser aus, wenn sie unmittelbar vor ihrem Tod fünf nur leicht deprimierende Jahre zusätzlich erlebte.

## 37. Erlebtes Wohlbefinden

1
Eine andere häufig gestellte Frage lautet: »Wie würden Sie, alles in allem, Ihr Leben gegenwärtig beurteilen? Würden Sie sagen, dass Sie sehr glücklich, ziemlich glücklich oder nicht besonders glücklich sind?« Diese Frage ist im General Social Survey in den Vereinigten Staaten enthalten, und ihre Korrelationen mit anderen Variablen deuten auf eine Mischung aus Zufriedenheit und Glückserleben hin. Ein reines Maß der Lebenszufriedenheit, das in den Gallup-Umfragen verwendet wird, ist die *Cantril Self-Anchoring Striving Scale*, auf der der Befragte seine gegenwärtige Lebenszufriedenheit auf einer Leiterskala einstuft, bei der null für die »schlimmste mögliche Lebenslage für Sie« steht und zehn für die »bestmögliche Lebenslage für Sie«. Die Formulierung legt nahe, dass Menschen ihr Urteil in dem verankern sollten, was sie für sich als möglich erachten, aber die empirischen Daten zeigen, dass die Menschen überall auf der Welt einen gemeinsamen Standard dafür haben, was ein gutes Leben ausmacht. Dies erklärt die außerordentlich hohe Korrelation (r = 0,84) zwischen dem Bruttoinlandsprodukt von Ländern und dem durchschnittlichen Score ihrer Bürger auf der Leiterskala. Angus Deaton, »Income, Health, and Well-Being Around the World: Evidence from the Gallup World Poll«, *Journal of Economic Perspectives* 22 (2008): S. 53–72.
2
Der Wirtschaftswissenschaftler war Alan Krueger von der Universität Princeton, der bekannt ist für seine innovativen Analysen ungewöhnlicher Daten. Die Psychologen waren David Schkade, ein Experte für Methodenlehre, Arthur Stone, ein Experte für Gesundheitspsychologie, Erfahrungsstichproben und die Methode des sogenannten »Ecological Momentary Assessment«, und der Sozialpsychologe Norbert Schwarz, der auch ein Experte für Erhebungsmethodik ist und selbst Experimente durchgeführt hat, die die Ergebnisse der Wohlbefindensforschung kritisieren, einschließlich des Experiments, bei dem ein Zehn-Cent-Stück, das auf einen Kopierer gelegt wurde, die anschließenden Angaben zur Lebenszufriedenheit beeinflusste.
3

Bei einigen Anwendungen liefert die Testperson auch physiologische Daten, etwa durch kontinuierliche Aufzeichnung des Herzschlags, gelegentliche Aufzeichnungen des Blutdrucks oder Speichelproben zur chemischen Analyse. Diese Methode wird *Ecological Momentary Assessment* genannt: Arthur A. Stone, Saul S. Shiffman und Marten W. DeVries, »Ecological Momentary Assessment Well-Being: The Foundations of Hedonic Psychology«, in Kahneman, Diener und Schwarz, *Well-Being*, S. 26–39.

4

Daniel Kahneman et al., »A Survey Method for Characterizing Daily Life Experience: The Day Reconstruction Method«, *Science* 306 (2004): S. 1776–1780. Daniel Kahneman und Alan B. Krueger, »Developments in the Measurement of Subjective Well-Being«, *Journal of Economic Perspectives* 20 (2006): S. 3–24.

5

Frühere Forschungen hatten belegt, dass Menschen Gefühle, die sie in einer vergangenen Situation hatten, »wieder erleben« können, wenn die Situation in hinlänglich anschaulichen Details abgerufen wird. Michael D. Robinson und Gerald L. Clore, »Belief and Feeling: Evidence for an Accessibility Model of Emotional Self-Report«, *Psychological Bulletin* 128 (2002): S. 934–960.

6

Alan B. Krueger (Hg.), *Measuring the Subjective Well-Being of Nations: National Accounts of Time Use and Well-Being*, Chicago 2009.

7

Ed Diener, »Most People Are Happy«, *Psychological Science* 7 (1996): S. 181–185.

8

Einige Jahre lang bin ich einer von mehreren Wissenschaftlern gewesen, die die Gallup Organization bei ihren Aktivitäten auf dem Gebiet der Glücksforschung berieten.

9

Daniel Kahneman und Angus Deaton, »High Income Improves Evaluation of Life but Not Emotional Well-Being«, *Proceedings of the National Academy of Sciences* 107 (2010): S. 16489–16493.

10

Dylan M. Smith, Kenneth M. Langa, Mohammed U. Kabeto und Peter Ubel, »Health, Wealth, and Happiness: Financial Resources Buffer Subjective Well-Being After the Onset of a Disability«, *Psychological Science* 16 (2005): S. 663–666.

11

Bei einem Vortrag auf einer TED-Konferenz im Februar 2010 erwähnte ich eine vorläufige Schätzung von 60 000 Dollar, die später korrigiert wurde.

12

Jordi Quoidbach, Elizabeth W. Dunn, K. V. Petrides und Moïra Mikolajczak, »Money Giveth, Money Taketh Away: The Dual Effect of Wealth on Happiness«, *Psychological Science* 21 (2010): S. 759–763.

## 38. Lebenszufriedenheit

1
Andrew E. Clark, Ed Diener und Yannis Georgellis, »Lags and Leads in Life Satisfaction: A Test of the Baseline Hypothesis«, Referat, vorgetragen auf der German Socio-Economic Panel Conference, Berlin 2001.

2
Daniel T. Gilbert und Timothy D. Wilson, »Why the Brain Talks to Itself: Sources of Error in Emotional Prediction«, *Philosophical Transactions of the Royal Society B* 364 (2009): S. 1335–1341.

3
Strack, Martin und Schwarz, »Priming and Communication«.

4
Über die ursprüngliche Studie berichtete Norbert Schwarz in seiner Doktorarbeit *Stimmung als Information. Untersuchungen zum Einfluss von Stimmungen auf die Bewertung des eigenen Lebens*, Heidelberg 1987. Sie wurde in zahlreichen Publikationen beschrieben, insbesondere von Norbert Schwarz und Fritz Strack, »Reports of Subjective Well-Being: Judgmental Processes and Their Methodological Implications«, in Kahneman, Diener und Schwarz, *Well-Being*, S. 61–84.

5
Die Studie wurde beschrieben in William G. Bowen and Derek Curtis Bok, *The Shape of the River*: *Long-Term Consequences of Considering Race in College and University Admissions*, Princeton 1998. Über einige der Untersuchungsergebnisse von Bowen und Bok berichteten Carol Nickerson, Norbert Schwarz und Ed Diener, »Financial Aspirations, Financial Success, and Overall Life Satisfaction: Who? and How?«, *Journal of Happiness Studies* 8 (2007): S. 467–515.

6
Alexander Astin, M. R. King und G. T. Richardson, »The American Freshman: National Norms for Fall 1976«, Cooperative Institutional Research Program of the American Council on Education der Universität von Kalifornien in Los Angeles, Graduate School of Education, Laboratory for Research in Higher Education, 1976.

7

Diese Ergebnisse wurden in einem Vortrag auf der Jahrestagung 2004 der American Economic Association vorgestellt. Daniel Kahneman, »Puzzles of Well-Being«, Tagungsreferat.

8

Die Frage, wie gut Menschen heute die Gefühle ihrer Nachkommen in hundert Jahren vorhersagen können, ist eindeutig für die politische Antwort auf den Klimawandel von Bedeutung, aber sie lässt sich nur indirekt untersuchen, was unsere Absicht war.

9

Als ich diese Frage stellte, warf ich zwei Begriffe in einen Topf, die nicht in einen Topf gehören: Glück und Lebenszufriedenheit sind nicht dasselbe. Lebenszufriedenheit bezieht sich auf die Gedanken und Gefühle, die einem einfallen, wenn man über sein Leben nachdenkt, was gelegentlich geschieht – auch in Umfragen zum Wohlbefinden. Glück beschreibt die Gefühle, die Menschen haben, wenn sie ihr normales Leben leben.

10

Meine Frau hat das allerdings nie eingestanden. Sie behauptet, dass nur Einwohner von Nordkalifornien glücklicher sind.

11

Asiatische Studenten berichteten im Allgemeinen eine geringere Lebenszufriedenheit, und asiatische Studenten stellten einen höheren Prozentsatz der Stichproben in Kalifornien als im Mittleren Westen. Bereinigt um diesen Unterschied, war die Lebenszufriedenheit in den beiden Regionen gleich.

12

Jing Xu und Norbert Schwarz haben herausgefunden, dass die Qualität des Autos (gemessen anhand seines Wertes in der Gebrauchtwagenliste) die Antwort der Eigentümer auf eine allgemeine Frage über ihre Freude an ihrem Fahrzeug und auch den Spaß der Fahrer bei Spritztouren vorhersagt. Aber die Qualität des Autos wirkt sich nicht auf die Stimmung von Menschen beim normalen Pendeln zwischen Wohnsitz und Arbeitsplatz aus. Norbert Schwarz, Daniel Kahneman und Jing Xu, »Global and Episodic Reports of Hedonic Experience«, in R. Belli, D. Alwin und F. Stafford (Hg.), *Using Calendar and Diary Methods in Life Course Research*, Newbury Park 2009, S. 157–174.

13

Diese Studie ist ausführlicher beschrieben in Kahneman, »Evaluation by Moments«.

14

Camille Wortman und Roxane C. Silver, »Coping with Irrevocable Loss, Cataclysms, Crises, and Catastrophes: Psychology in Action«, American Psychological Association, *Master Lecture Series* 6 (1987): S. 189–235.

15

Dylan Smith et al., »Misremembering Colostomies? Former Patients Give Lower Utility Ratings than Do Current Patients«, *Health Psychology* 25 (2006): S. 688–695. George Loewenstein und Peter A. Ubel, »Hedonic Adaptation and the Role of Decision and Experience Utility in Public Policy«, *Journal of Public Economics* 92 (2008): S. 1795–1810.

16

Daniel Gilbert und Timothy D. Wilson, »Miswanting: Some Problems in Affective Forecasting«, in *Feeling and Thinking: The Role of Affect in Social Cognition*, hg. von Joseph P. Forgas, New York 2000, S. 178–197.

## Schlusswort

1
Paul Dolan und Daniel Kahneman, »Interpretations of Utility and Their Implications for the Valuation of Health«, *Economic Journal* 118 (2008): S. 215–234. Loewenstein und Ubel, »Hedonic Adaptation and the Role of Decision and Experience Utility in Public Policy«.

2
Vor allem in Großbritannien hat man besonders schnelle Fortschritte gemacht; hier ist die Verwendung von Maßen des Wohlbefindens jetzt offizielle Regierungspolitik. Diese Fortschritte verdanken sich zu einem großen Teil der starken Resonanz von Lord Richard Layards Buch *Die glückliche Gesellschaft: Kurswechsel für Politik und Wirtschaft (Happiness: Lessons from a New Science)*, das 2005 veröffentlicht wurde. Layard gehört zu den bekannten Ökonomen und Sozialwissenschaftlern, die sich der Erforschung des Wohlbefindens und seiner Folgen verschrieben haben. Weitere wichtige Quellen sind: Derek Bok, *The Politics of Happiness: What Government Can Learn from the New Research on Well-Being*, Princeton 2010. Ed Diener, Richard Lucus, Ulrich Schmimmack und John F. Helliwell, *Well-Being for Public Policy*, New York 2009. Alan B. Krueger (Hg.), *Measuring the Subjective Well-Being of Nations: National Account of Time Use and Well-Being*, Chicago 2009. Joseph E. Stiglitz, Amartya Sen und Jean-Paul Fitoussi, *Report of the Commission on the Measurement of Economic Performance and Social Progress*. Paul Dolan, Richard Layard und Robert Metcalfe, *Measuring Subjective Well-being for Public Policy: Recommendations on Measures*, London: Office for National Statistics, 2011.

3
Die Sichtweise des menschlichen Intellekts, die Dan Ariely in *Denken hilft zwar, nützt aber nichts: Warum wir immer wieder unvernünftige Entscheidungen treffen (Predictably Irrational: The Hidden Forces That Shape Our Decisions)*, München 2008, vorgelegt hat, unterscheidet sich nicht allzu sehr von meiner Konzeption, aber wir verwenden den Begriff »irrational« unterschiedlich.

4

Gary S. Becker und Kevin M. Murphy, »A Theory of Rational Addiction«, *Journal of Political Economics* 96 (1988): S. 675–700.

5

Richard H. Thaler und Cass R. Sunstein, *Nudge: Wie man kluge Entscheidungen anstößt (Nudge. Improving Decisions about Health, Wealth and Happiness)*, Berlin 2009.

6

Atul Gawande, *The Checklist Manifesto: How to Get Things Right*, New York 2009. Daniel Kahneman, Dan Lovallo und Oliver Sibony, »The Big Idea: Before You Make That Big Decision …«, *Harvard Business Review* 89 (2011): S. 50–60.

7

Chip Heath, Richard P. Larrick und Joshua Klayman, »Cognitive Repairs: How Organizational Practices Can Compensate for Individual Shortcomings«, *Research in Organizational Behavior* 20 (1998): S. 1–37.

## ANHANG Urteile unter Unsicherheit

1
Dieser Artikel erschien erstmals in *Science*, Bd. 185, 1974. Die Forschungsarbeiten wurden unterstützt von der Advanced Research Projects Agency des US-Verteidigungsministeriums und laut Vertrag N00014-73-C-0438 vom Oregon Research Institute, Eugene, überwacht. Außerdem wurden diesen Forschungen von der Forschungs- und Entwicklungsabteilung der Hebräischen Universität in Jerusalem gefördert.

2
D. Kahneman und A. Tversky, »On the Psychology of Prediction«, *Psychological Review* 80 (1973), S. 237–251.

3
Ebd.

4
Ebd.

5
D. Kahneman und A. Tversky, »Subjective Probability: A Judgment of Representativeness«, *Cognitive Psychology* 3 (1972): S. 430–454.

6
Ebd.

7
W. Edwards, »Conservatism in Human Information Processing«, in *Formal Representation of Human Judgment*, hg. von B. Kleinmuntz, New York 1968, S. 17–52.

8
Kahneman und Tversky, »Subjective Probability«.

9
A. Tversky und D. Kahneman, »Belief in the Law of Small Numbers«, *Psychological Bulletin* 76 (1971): S. 105–110.

10
Kahneman und Tversky, »On the Psychology of Prediction«.

11
Ebd.

12
Ebd.

13
Ebd.
14
A. Tversky und D. Kahneman, »Availability: A Heuristic for Judging Frequency and Probability«, *Cognitive Psychology* 5 (1973): S. 207–232.
15
Ebd.
16
R. C. Galbraith und B. J. Underwood, »Perceived Frequency of Concrete and Abstract Words«, *Memory & Cognition* 1 (1973): S. 56–60.
17
Tversky und Kahneman, »Availability«.
18
L. J. Chapman und J. P. Chapman, »Genesis of Popular but Erroneous Psychodiagnostic Observations«, *Journal of Abnormal Psychology* 73 (1967): S. 193–204. L. J. Chapman und J. P. Chapman, »Illusory Correlation as an Obstacle to the Use of Valid Psychodiagnostic Signs«, *Journal of Abnormal Psychology* 74 (1969): S. 271–280.
19
P. Slovic und S. Lichtenstein, »Comparison of Bayesian and Regression Approaches to the Study of Information Processing in Judgment«, *Organizational Behavior & Human Performance* 6 (1971): S. 649–744.
20
M. Bar-Hillel, »On the Subjective Probability of Compound Events«, *Organizational Behavior & Human Performance* 9 (1973): S. 396–406.
21
J. Cohen, E. I. Chesnick und D. Haran, »A Confirmation of the Inertial-Ψ Effect in Sequential Choice and Decision«, *British Journal of Psychology* 63 (1972): S. 41–46.
22
M. Alpert und H. Raiffa, unveröffentlichtes Manuskript; C. A. Stael von Holstein, »Two Techniques for Assessment of Subjective Probability Distributions: An Experimental Study«, *Acta Psychologica* 35 (1971): S. 478–494. R. L. Winkler, »The Assessment of Prior Distributions in Bayesian Analysis«, *Journal of the American Statistical Association* 62 (1967): S. 776–800.
23

Kahneman und Tversky, »Subjective Probability«. Tversky und Kahneman, »Availability«.

24

Kahneman und Tversky, »On the Psychology of Prediction«. Tversky und Kahneman, »Belief in the Law of Small Numbers«.

25

L. J. Savage, *The Foundations of Statistics*, New York 1954.

26

Ebd. und B. de Finetti, »Probability: Interpretations«, in *International Encyclopedia of the Social Sciences*, hg. von D. E. Sills, Band 12, New York 1968, S. 496–505.

## Entscheidungen, Werte und Frames

1
Dieser Beitrag geht auf ein Referat zurück, das auf der Jahrestagung der American Psychological Association im August 1983 mit dem Distinguished Scientific Contributions Award ausgezeichnet wurde. Er wurde gefördert durch einen Zuschuss des U.S. Office of Naval Research. Erstveröffentlichung in *American Psychologist* 34, 1984.

2
D. Bernoulli, »Exposition of a New Theory on the Measurement of Risk«, *Econometrica* 22 (1954 [1738]): S. 23–36.

3
D. Kahneman und A. Tversky, »Prospect Theory: An Analysis of Decision under Risk«, *Econometrica* 47 (1979): S. 263–291.

4
P. C. Fishburn und G. A. Kochenberger, »Two-Piece von Neumann – Morgenstern Utility Functions«, *Decision Sciences* 10 (1979): S. 503–518. J. C. Hershey und P. J. H. Schoemaker, »Risk Taking and Problem Context in the Domain of Losses: An Expected-Utility Analysis«, *Journal of Risk and Insurance* 47 (1980): S. 111–132. J. W. Payne, D. J. Laughhunn und R. Crum, »Translation of Gambles and Aspiration Level Effects in Risky Choice Behavior«, *Management Science* 26 (1980): S. 1039–1060. P. Slovic, B. Fischhoff und S. Lichtenstein, »Response Mode, Framing, and Information-Processing Effects in Risk Assessment«, in *New Directions for Methodology of Social and Behavioral Science: Question Framing and Response Consistency*, hg. v. R. Hogarth, San Francisco 1982, S. 21–36.

5
S. E. Erakar und H. C. Sox, »Assessment of Patients' Preferences for Therapeutic Outcomes«, *Medical Decision Making* 1 (1981): S. 29–39.

6
B. Fischhoff, »Predicting Frames«, *Journal of Experimental Psychology: Learning, Memory and Cognition* 9 (1983): S. 103–116. A. Tversky, »On the Elicitation of Preferences: Descriptive and Prescriptive Considerations«, in *Conflicting Objectives in Decisions*, hg. v. D. Bell, R. L. Kenney und H. Raiffa, New York 1977, S. 209–222. A. Tversky und D. Kahneman, »The

Framing of Decisions and the Psychology of Choice«, *Science* 211 (1981): S. 453–458.

7

J. von Neumann und O. Morgenstern, *Theory of Games and Economic Behavior*, 2. Aufl., Princeton 1947.

8

M. Allais und O. Hagen (Hg.), *Expected Utility Hypotheses and the Allais Paradox*, Hingham 1979.

9

A. Tversky und D. Kahneman, »The Framing of Decisions and the Psychology of Choice«.

10

R. Schlaifer, *Probability and Statistics for Business Decisions*, New York 1959.

11

B. Fischhoff, P. Slovic und S. *Lichtenstein*, »Knowing What You Want: Measuring Labile Values«, in *Cognitive Processes in Choice and Decision Behavior*, hg. v. T. Wallsten, Hillsdale 1980, S. 117–141.

12

D. Kahneman und A. Tversky, »Prospect Theory: An Analysis of Decision under Risk«.

13

P. Slovic, B. Fischhoff und S. Lichtenstein, »Response Mode, Framing, and Information-Processing Effects in Risk Assessment«.

14

B. McNeil, S. Pauker, H. Sox Jr. und A. Tversky, »On the Elicitation of Preferences for Alternative Therapies«, *New England Journal of Medicine* 306 (1982): S. 1259–1262.

15

R. Thaler, »Toward a Positive Theory of Consumer Choice«, *Journal of Economic Behavior and Organization* 1 (1980): S. 39–60.

16

H. H. Clark und E. V. Clark. *Psychology and Language*, New York 1977.

17

R. Thaler, »Toward a Positive Theory of Consumer Choice«. R. Thaler, »Using Mental Accounting in a Theory of Consumer Behavior«, *Marketing Science* 4 (1985): S. 199–214.

18
L. J. Savage, *The Foundation of Statistics*, New York 1954.
19
R. Thaler, »Toward a Positive Theory of Consumer Choice«.
20
J. W. Pratt, D. Wise und R. Zeckhauser, »Price Differences in Almost Competitive Markets«, *Quarterly Journal of Economics* 93 (1979): S. 189–211.
21
D. Kahneman, »The Simulation Heuristic«, in *Judgment Under Uncertainty: Heuristics and Biases*, hg. v. D. Kahneman, P. Slovic und A. Tversky, New York 1982, S. 201–208.
22
R. Thaler, »Toward a Positive Theory of Consumer Choice«.
23
R. Gregory, »Measures of Consumer's Surplus: Reasons for the Disparity in Observed Values«, unveröffentlichtes Manuskript, Keene State College, 1983. J. Hammack und G. M. Brown Jr., *Waterfowl and Wetlands: Toward Bioeconomic Analysis*, Baltimore 1974. J. Knetsch und J. Sinden, »Willingness to Pay and Compensation Demanded: Experimental Evidence of an Unexpected Disparity in Measures of Value«, *Quarterly Journal of Economics* 99 (1984): S. 507–521.
24
P. Slovic, B. Fischhoff und S. Lichtenstein, »Response Mode, Framing, and Information-Processing Effects in Risk Assessment«.
25
P. J. H. Schoemaker und H. C. Kunreuther, »An Experimental Study of Insurance Decisions«, *Journal of Risk and Insurance* 46 (1979): S. 603–618.
26
J. C. Hershey und P. J. H. Schoemaker, »Risk Taking and Problem Context in the Domain of Losses: An Expected-Utility Analysis«.
27
R. Thaler, »Toward a Positive Theory of Consumer Choice«.
28
J. G. March, »Bounded Rationality, Ambiguity, and the Engineering of Choice«, *Bell Journal of Economics* 9 (1978): S. 587–608.

29
P. Brickman und D. T. Campbell, »Hedonic Relativism and Planning the Good Society«, in *Adaptation Level Theory: A Symposium*, hg. v. M. H. Appley, New York 1971, S. 287–302.

# *Sachregister*

3-D-Heuristik

Above-Average-Effekt
Abrufflüssigkeit
Abrufleichtigkeit siehe Abrufflüssigkeit
Additions-Aufgabe
Affektheuristik
Agenten, rationale siehe Modell des rationalen Agenten
Aktienmarkt
Aktivierung, assoziative
Aktivierungsausbreitung
Aktivierungszustand, körperlicher
Alar-Zwischenfall
Algorithmus
Allais-Paradoxon
Ambiguität
Amygdala
Ankereffekt
Ankerheuristik siehe Ankereffekt
Ankerungsindex
Anomalien
Anpassungsniveau
Anstrengung
Apgar-Score
Aphorismen
Armut
*arousal* siehe Aktivierungszustand, körperlicher
Artefakte
Assoziationsmaschine
– paar

Aufmerksamkeitssteuerung
Aussagekraft, prognostische

Außensicht
Axiome

Bankkassiererin-Problem
Basisprognosen
– raten

Basisratenfehler
Basisraten, kausale
– , nonkausale
– , statistische

Bayessche Statistik
Beanspruchung, kognitive
Begabung siehe Talent
Beharrlichkeit, irrationale
Beinbruch-Regel
Belohnungen
Bernoullis Theorie
Besitztumseffekt siehe Endowment-Effekt
Bestrafungen
Between-Subjects-Experiment
Bewertbarkeitshypothese
Bewertungen, einfache
– , elementare
– , gemeinsame

Blindheit, theorieinduzierte
Buchführung, mentale

*Cantril Self-Anchoring Striving Scale*
*Cognitive Reflection Test* (CRT)
*Cortex cingularis anterior*
Cortex, präfrontaler

Darmspiegelung
*Day Reconstruction Method* (DRM)
Denken, langsames

Denken, schnelles
Depression
Dispositionseffekt
DNA-Beweis
*Draw-a-Person-Test* (DAP)

*Econs* und *Humans*
Effekt der bloßen Darbietung siehe Mere-Exposure-Effekt
Effekt, ideomotorischer
Ego-Depletion
Einrahmungseffekt siehe Framing-Effekt
Eintrittskarten
Elektroschocks
Endowment-Effekt
Entscheidungsarchitektur
Entscheidungsfindung
– unter Unsicherheit
– , demokratische
– , intuitive
– , rationale

Entscheidungsgewichte
– nutzen

Ereignisse, konjunkte und disjunkte
– , seltene
– , ungewöhnliche

Erfahrung, optimale siehe Flow
Erfahrungsnutzen
Erfahrungsstichprobe
Ergebnisfehler
Erinnerungen
Erkenntnisillusion
Ersetzung
Erwartungsnutzentheorie
Erwartungsprinzip
Expertenwissen siehe Expertise

Expertise

Fechnersches Gesetz
Feedback
Fehler, dekorrelierte
Fehlschluss aus versunkenen Kosten siehe Kosten, versunkene
Feuerwehrmänner
Finanzexperten
Florida-Effekt
Flow
Fokalismus
Fokussierungs-Illusion
Formeln
Fragen, heuristische
Framing-Effekt
Framing, emotionales
– , enges
– , weites

Freizeit
Fremde
Funktionen, logarithmische

Gallup World Poll
*Gallup-Healthways Well-Being Index*
Gedächtnis, assoziatives
Gehirn
Gehirnscan
Gesetz der großen Zahlen
Gesetz der kleinen Zahlen
Gewichtung der Dauer
Glücksgefühl
Glücksrad
Glücksspiele
Golf
Gorilla-Experiment
Grundgesamtheit

Grundhäufigkeit siehe Basisraten
Gutschein-Experiment

Halo-Effekt
Häufigkeitsrepräsentation
Hedonimeter
Heiraten
Heuristik (Definition)
Hilfe-Experiment
Hinweisreiz
Hirnaktivität
Hirnareal
Hirnforschung
Hirnscan siehe Gehirnscan
Hirnschaden
Höchststand-Ende-Regel
*Humans* und *Econs* siehe *Econs* und *Humans*
Hybris-Hypothese

Ideen, autoritäre
Illusion der Gültigkeit
– der Kausalität siehe Kausalitätsillusion
– der Kompetenz siehe Kompetenzillusion
– der Kontrolle
– der Prognosegültigkeit
– des Verstehens

*illusion of skill* siehe Kompetenzillusion
Illusion, kognitive
Illusionen von Experten
– von Unternehmern

Immunsystem, psychologisches
Impfstoffe
Indifferenzkurve
Innensicht
Intensitätsabstimmung
Intensitätsskala

Interpretation, kausale
Invarianzkriterium
Investitionen
Investoren

Julie-Problem

Kaffeetassen-Experimente
Kalte-Hand-Experiment
Kältedrucktest siehe Kalte-Hand-Experiment
Kategorien
Kausalität, physikalische und intentionale
Kausalitätsillusion
Keks-Experiment
Kognitionsmechanismus
Kohärenz, assoziative
Kolostomie-Patienten
Kompetenzillusion
Konfidenzintervall
Konjunktionsfehlschluss
Konkurrenzvernachlässigung
Konten, mentale
Kontrolle, exekutive
– , kognitive
Kontrollgruppe
Korrelation, illusorische
Korrelationskoeffizient
Kosten-Nutzen-Abwägung, tabuisierte siehe *taboo trade-off*
Kosten, versunkene (irreversible)
Kraftstoffverbrauch
Krebs

Lächeln
Lady-Macbeth-Effekt
Lebenszufriedenheit
Leichtigkeit, kognitive
Linda-Problem

Lotterien siehe Glücksspiele
Love-Canal-Affäre

Marsch der Geschichte
Maschinerie, assoziative
Medizin
Mere-Exposure-Effekt
Michigan-Detroit-Problem
Misswollen
Modell der intuitiven Entscheidungsfindung
– der wiedererkennungsgesteuerten Entscheidung siehe RPD-Modell
– des rationalen Agenten

Möglichkeitseffekt
Moses-Illusion
MPG-Illusion
Müller-Lyer-Illusion
Münze-und-Kopierer-Experiment

*Naturalistic Decision Making* (NDM) siehe Naturalistische Entscheidungsfindung
Naturalistische Entscheidungsfindung
Negativitätsdominanz
Nenner-Vernachlässigung
Neue Erwartungstheorie
Neuroökonomie
Normen
Normentheorie
Nutzentheorie

Oakland A ’s
Optimismus
Optimismus-Verzerrung
Organspende

Paternalismus, libertärer
Perzentile
Planungsfehlschluss

Prä-mortem-Methode
Prädiktoren
Präferenztheorie
– umkehr

Priming-Effekt
Prinzip unabhängiger Urteile
Problem der Asiatischen Krankheit
*Prospect Theory* siehe Neue Erwartungstheorie
Prototypen
Prozentrang siehe Perzentile
Psychophysik
Pupillen
Pupillometrie, kognitive

Rahmen, Rahmung siehe Framing-Effekt
Rationalität
Rationalitätsindex
Rationalitätsregeln siehe Axiome
RPD-Modell
Referenzklassenprognose
Referenzpunkt
Reframing
Regression zum Mittelwert
Regression, multiple
Regressionseffekt
Rekognitionsheuristik
*Remote Association Test* (RAT)
Repräsentativität
Repräsentativitätsheuristik
– urteil

Risikoaversion
– bewertung
– freude
– strategien

Rückschaufehler

Sankt-Petersburg-Paradoxon
Schach
Schläger-Ball-Problem
Schlussfolgerungen, voreilige
Schmerzen
Schrotflinte, mentale
Selbst, erinnerndes
– , erlebendes

Selbste
Selbstkontrolle
– kritik
– überschätzung

Service-Problem
Sicherheitseffekt
Situationen, kausale
Spielerfehlschluss
Status quo
Steigerung des Einsatzes
Stellvertreter-Problem
Stereotype
Steve, der Bibliothekar
Stichprobenvarianz
Stimmungsheuristik
Stirnrunzeln
Stockpicker, Stock-Picking
Strafen siehe Bestrafungen
Suggestion
Syllogismen
System
System
System, assoziatives

*taboo trade-off*
Tadel
Talent

*task sets*
Täuschungen
Taxi-Problem
Termini, assoziative
Terrorismus
Teststrategie, positive
Theorie der Entscheidungsfindung
Theorie des erwarteten Nutzens siehe Erwartungsnutzentheorie
Tom-W.-Problem
Totalverlust-Effekt

U-Index
Überraschung
Überzeugtsein
Umgebungen niedriger Validität
Unternehmer
Urlaub
Urteile, durationsgewichtete
Urteilsheuristik
Urteilssprünge siehe voreilige Schlussfolgerungen
– theorie

Variablen, summenähnliche
Venn-Diagramme
Verantwortung
Verantwortungsdiffusion
Verarbeitungsflüssigkeit siehe Leichtigkeit, kognitive
Verfügbarkeit
Verfügbarkeitseffekte
– fehler
– heuristik
– kaskaden
– urteile

Verhaltensökonomik
Verknüpfung, assoziative
Verlustaversion

Verlustaversionskoeffizient
Verlustaversionsrate
Vernachlässigung der Dauer
Vertrautheit
Verzerrung, kognitive
– , narrative

Viergeteiltes Muster
Volkswirtschaftslehre
Vorhersage, intuitive
Vorhersagefehler, affektiver
Vorhersagevariablen siehe Prädiktoren
Vorstandschef

Wagniskapitalgesellschaften
Wahlen auf Beschreibungsgrundlage
– auf Erfahrungsgrundlage

Wahrheitsillusionen
Wahrscheinlichkeitsempfindlichkeit
– urteil
– vernachlässigung
– verteilung

Webersches Gesetz
Wein
Weniger-ist-mehr-Muster
Wertpapierdepot
Wiederholung
Wirtschaftswissenschaft
Within-Subject-Experiment
Wohlbefinden
Wollbarkeit
WYSIATI *(What you see is all there is)*

Yale-Prüfungs-Problem

Zeitdruck

*zero-validity environment* siehe Aussagekraft, prognostische
Zufallsanker
– prozess

Zufriedenheit
Zweifel

# *Personenregister*

Ajzen, Icek
Allais, Maurice
*American Economic Review*
Apgar, Virginia
Ariely, Dan
Arrow, Kenneth
Asch, Solomon
Ashenfelter, Orley
Åstebro, Thomas
*The Atlantic*
Auerbach, Red

Barber, Brad
– »Wertpapiergeschäfte bedrohen Ihr Vermögen«

Bargh, John
Baumeister, Roy
Bayes, Thomas
Bazerman, Max
– *Judgement in Managerial Decision Making*

Beane, Billy
Beatty, Jackson
Becker, Gary
Benartzi, Shlomo
Bentham, Jeremy
– *Einführung in die Prinzipien der Moral und der Gesetzgebung*

Berlin, Isaiah
– »Der Igel und der Fuchs«

Bernoulli, Daniel
Bernoulli, Nicholas

Beyth, Ruth
Borg, Björn
Borgida, Eugene
Bradlee, Ben
Brockman, John
Bush, George W.

Cabanac, Michael
Camerer, Colin
Carroll, Lewis
– »Jabberwocky«

Chabris, Christopher
– *Der unsichtbare Gorilla*

Clark, Andrew
Clinton, Bill
Coelho, Martha
Cohen, David
Cohn, Beruria
Collins, Jim
– (und J. Porras) *Immer erfolgreich*

Csikszentmihalyi, Mihaly

Damasio, Antonio
Dawes, Robyn
– »The Robust Beauty of Improper Linear Models in Decision Making«

Deaton, Angus
Diener, Ed
Dosi, Giovanni

*Econometrica*
*Edge*
Edgeworth, Francis
Epley, Nick
Epstein, Seymour

Erev, Ido

Fechner, Gustav Theodor
Feller, William
Fischhoff, Baruch
Flyvbjerg, Bent
*Fortune*
Fox, Craig
Fox, Seymour
Frederick, Shane
Freedman, David
Frey, Bruno
Friedman, Milton
– *Chancen, die ich meine*

Galinsky, Adam
Galton, Francis
– »Regression towards Mediocrity in Hereditary Stature«

Gawande, Atul
– *A Checklist Manifesto*

Georgellis, Yannis
Gibbs, Lois
Gigerenzer, Gerd
Gilbert, Daniel
– *Ins Glück stolpern*
– »Wie mentale Systeme glauben«

Gilovich, Tom
Gladwell, Malcolm
– *Blink*

Gottman, John
Gould, Stephen Jay
Grether, David
Guthrie, Chris

Haidt, Jonathan
Harding, Warren G.
Heider, Fritz
Hertwig, Ralph
Hess, Eckhard
Hitler, Adolf
Hogarth, Robin
Hsee, Christopher
Hume, David
– *Eine Untersuchung über den menschlichen Verstand*

Jacoby, Larry
– »Über Nacht berühmt werden«

Jencks, Christopher

Kaye, Danny
Klein, Gary
– (und D. Kahneman) »Conditions for Intuitive Expertise: A Failure to Disagree«
– *Natürliche Entscheidungsprozesse*

Knetsch, Jack
Krueger, Alan
Kunreuther, Howard
Kuran, Timur

Larrick, Richard
– (und J. Soll) »The MPG Illusion«

Larson, Gary
– *The Far Side*

Layard, Richard
LeBoeuf, Robyn
Lewis, Michael
– *Moneyball*

Lichtenstein, Sarah
List, John
Lovallo, Dan

Malkiel, Burton
– *Börsenerfolg ist kein Zufall*

Malmendier, Ulrike
Mao Tsetung
Markowitz, Harry
McFarland, Cathy
Mednick, Sarnoff
Meehl, Paul
– *Clinical vs. Statistical Prediction*

Michotte, Albert
Miller, Dale
Mischel, Walter
Morgenstern, Oskar
Mullainathan, Sendhil
Mussweiler, Thomas

Neumann, John von
*The New York Times*
Nisbett, Richard
Nixon, Richard

Obama, Barack
Odean, Terry
– »Wertpapiergeschäfte bedrohen Ihr Vermögen«

Oppenheimer, Danny
– »Consequences of Erudite Vernacular Utilized Irrespective of Necessity: Problems with Using Long Words Needlessly«

Pawlow, Ivan
Plott, Charles
Pólya, George

– *Schule des Denkens*

Pope, Devin
Porras, Jerrry
– (und J. Collins) *Immer erfolgreich*

*Proceedings of the National Academy of Sciences*

Rabin, Matthew
Reagan, Ronald
Redelmeier, Don
Rice, Condoleezza
Rosett, Richard
Rosenzweig, Philip
– *Der Halo-Effekt*

Rozin, Paul
Rumsfeld, Donald

Saddam Hussein
Samuelson, Paul
Savage, Jimmie
*Science*
*Scientific American*
Schelling, Thomas
– *Choice and Consequence*

Schkade, David
Schwarz, Norbert
Schweitzer, Maurice
Seligman, Martin
Shafir, Eldar
Simmel, Mary-Ann
Simon, Herbert
Simons, Daniel
– (und C. Chabris) *Der unsichtbare Gorilla*

Simpson, O. J.

Slovic, Paul
Smith, Vernon
Soll, Jack
– (und R. Larrick) »The MPG Illusion«

Spinoza, Baruch de
*Sports Illustrated*
Stalin, Josef
Stanovich, Keith
– *Rationality and the Reflective Mind*

Steiger, James H.
Stone, Arthur
Strack, Fritz
Streep, Meryl
Sunstein, Cass
– (und R. Thaler) *Nudge*

Surowiecki, James
– *Die Weisheit der Vielen*

Taleb, Nassim
– *Der Schwarze Schwan*

Tate, Geoffrey
Tenet, George
Tetlock, Philip
– *Expert Political Judgment: How Good is it? How Can We Know?*

Thaler, Richard
– (und C. Sunstein) *Nudge*

Thomas, Lewis
*Todesboden*
Todorov, Alex
*La Traviata* (Verdi)
Truman, Harry
Tversky, Amos passim

– (und D. Kahneman) »Glaube an das Gesetz der kleinen Zahlen«
– (undKahneman) »Judgement Under Uncertainty: Heuristics and Biases«
– (et al.) *Mathematical Psychology*
– (undKahneman) »Prospect Theory: An Analysis of Decision Under Risk«

Vallone, Robert
Vohs, Kathleen

Wainer, Howard
*The Washington Post*
Weber, Ernst Heinrich
West, Richard
Wilson, Timothy
– *Strangers to Ourselves*

Woods, Tiger

Xu, Jing

Zajonc, Robert
Zamir, Eyal
Zeiler, Kathryn
Zwerling, Harris

Die amerikanische Originalausgabe erschien 2011 unter dem Titel
*Thinking, fast and slow* bei Farrar, Straus and Giroux, New York.

Die Rechte für die Abbildungen im Innenteil liegen bei:
Elyssa Match/Paul Ekman Group (S. 31), Royal Society (S. 78),
© 1990 by Roger N. Shepard. Mit freundlicher Genehmigung
Henry Holt and Company, LLC (S. 130),
Science Magazine (S. 369).

Erste Auflage

Copyright © 2011 by Daniel Kahneman
Copyright © der deutschsprachigen Ausgabe 2012 by Siedler Verlag, München,
in der Verlagsgruppe Random House GmbH

Umschlaggestaltung: Rothfos + Gabler, Hamburg
Lektorat: Nico Schröder, Hamburg
Satz: Ditta Ahmadi, Berlin

eISBN 978-3-641-09374-7

www.siedler-verlag.de
www.randomhouse.de

a
Merkmale, die in Teil 4 ausführlich dargestellt werden.

Daniel Kahneman

# SCHNELLES DENKEN, LANGSAMES DENKEN

Aus dem amerikanischen Englisch von
Thorsten Schmidt

Siedler

This file was downloaded from Z-Library project

*Your gateway to knowledge and culture. Accessible for everyone.*

z-library.sk z-lib.gs z-lib.fm go-to-library.sk

Official Telegram channel

Z-Access

https://wikipedia.org/wiki/Z-Library